# Exploring the mechanisms driving temporally dynamic spatial frequency processing in the primary visual cortex of mice

Ahmad Elsayed
Charlottesville, VA

BSc. Psychology, George Mason University, 2017

A Dissertation (*or Thesis*) presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy or Master of Arts or Master of Science or Master of Fine Arts

Department of Psychology

University of Virginia
November 2023

Jianhua “JC” Cang

Chad Daniel Meliza

# 1. Abstract

Neurons of the primary visual cortex (V1) process spatial frequency information in a sequential fashion whereby low spatial frequencies (SFs) are processed prior to high SFs. This phenomenon, known as coarse-to-fine processing, has been well documented across several mammalian species and is involved in several important visual computations such as feature-linking, core object recognition, and the efficient coding of natural scenes. Despite the importance of coarse-to-fine processing in visual computation, a complete description of the mechanisms behind this phenomenon has yet to be established. Prior studies have questioned whether coarse-to-fine processing stems from the primary feedforward visual pathway or from local cortical interactions and feedback from higher visual areas. This study attempts to investigate this question by determining whether a purely feedforward model can produce coarse-to-fine-like responses and whether such a model is biologically plausible. I have found that a biologically plausible feedforward model can only produce coarse-to-fine-like responses that are limited to the lower band of SFs; that is, the model responded to progressively higher SFs with progressively longer time lags but did not respond to SFs beyond a certain value. This result indicates that a purely feedforward model cannot explain the high SF portion of the coarse-to-fine response but may explain the low SF portion. However, because recorded V1 neurons still show prominent responses to high SF, high SF information is likely present in the feedforward pathway and is potentially enhanced via intracortical interactions or feedback from higher visual areas.

# 2. Introduction

Generally, the global features of visual scenes are perceived prior to the local details of a visual stimulus. For example, human participants can be asked to discriminate between the global and local identity of a figure resembling a large letter which is composed of a different, smaller letter. In such a task, the identity of the large letter will typically interfere with the participants' ability to identify the identity of the small letters; however, the large letters are typically identified with no hesitation. The fact that attention cannot be efficiently diverted from global features during this task suggests that global processing is a necessary and early stage of visual perception (Navon, D., 1977).

Neurons in the primary visual cortex of several animals physiologically respond to coarse and fine visual features in a similar temporal fashion to how human participants behaviorally respond to global and local visual features. In experiments that explore this similarity, neurons are extracellularly recorded as animals are shown sinusoidal grating stimuli on a computer monitor. Experimenters vary the spatial frequencies (SFs) of these stimuli to build a picture of how SF preference changes over time. Studies of this nature have shown that in monkeys (Bredfeldt, C. E. & Ringach, D. L., 2002.), cats (Tanaka, H. & Sawada, R., 2022), and mice (Skyberg, R., Tanabe, S., Chen, H. & Cang, J., 2022), neurons of the primary visual cortex shift their SF preference over short timescales from low SF in the early stage of the response to high SF in the later stage. Results from my study also show that coarse-to-fine processing is a continuous and smooth process, with responses to progressively higher SFs occurring at progressively longer time lags even if the bandwidth of SF preference is limited to lower SFs.

Understanding coarse-to-fine processing is important to understanding many elements of visual perception. For instance, coarse-to-fine processing has been implicated as a mechanism facilitating feature-linking (Sceniak, M. P., Ringach, D. L., Hawken, M. J. & Shapley, R., 1999). Coarse-to-fine processing may also be a mechanism facilitating core object recognition (Sceniak, M. P., Ringach, D. L., Hawken, M. J. & Shapley, R., 1999). In concert with core object recognition, coarse-to-fine processing helps to reduce ambiguities within the local space of visual analysis (Purushothaman, G., Chen, X., Yampolsky, D. & Casagrande, V. A., 2014). For instance, at an immediate glance a reader may not be able to differentiate between the words “cat” and “cut” without allocating attention to the middle letters. Furthermore, coarse-to-fine processing provides a computational advantage to V1 for processing natural scenes. When neurons in V1 respond to natural scenes, their responses become decorrelated as SF preference shifts from low SF to high SF, eliminating redundancies in the population response (Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M. & Harris, K. D., 2019; Skyberg, R., Tanabe, S., Chen, H. & Cang, J., 2022).

The relevance of coarse-to-fine processing to such a wide array of visual functions results in a strong need for a mechanistic explanation behind the phenomenon. Previous research has focused on determining whether coarse-to-fine processing occurs because of hierarchical feedforward local inhibition, or feedback mechanisms; however, this question has not yet had a definitive answer. Some of the earliest studies that have attempted to answer this question claim that coarse-to-fine processing is facilitated by cortical inhibition because coarse-to-fine processing does not occur in upstream stages of visual processing such as the dorsal lateral geniculate nucleus (dLGN) (Bredfeldt, C. E. & Ringach, D. L., 2002). However, other early and more recent studies have claimed that coarse-to-fine processing is indeed present in dLGN and even retinal ganglion cells (RGCs) (Suder, K., Funke, K., Zhao, Y., Kerscher, N., Wennekers, T. & Wörgötter, F., 2002; Einevoll, G. T., Jurkus, P. & Heggelund, P., 2011; Purushothaman, G., Chen, X., Yampolsky, D. & Casagrande, V. A., 2014), suggesting that coarse-to-fine processing occurs via feedforward mechanisms.

In this study, I have attempted to determine whether a purely feedforward mechanism can explain why responses shift their SF frequency from low SF to high SF in neurons of the V1 of mice. Feedforward mechanisms produce linear spatiotemporal receptive fields (STRFs). These are defined as a set of filters in space and time that describe the features of the visual field that maximally elicit excitatory or suppressive activity in visual neurons. Taking the linear combination of a visual input with the STRF can produce an estimate of the expected response of visual neurons whose activity relies on feedforward mechanisms. Thus, I first set out to determine if there exist certain STRF dynamics which can reproduce coarse-to-fine-like responses using a linear-nonlinear Poisson (LNP) cascade model. Modeling allowed me to parameterize certain features of the STRF and analyze the relative importance of these parameters to producing the coarse-to-fine response. I then mapped STRFs from neurons recorded from the V1 of mice using a stimulus which could reliably elicit coarse-to-fine responses to get a ground-truth picture of how V1 STRFs change over time. Finally, I compared the similarity between ground-truth STRFs and simulated STRFs used in the model to provide an inference of whether the model is biologically plausible.

Results from the LNP model showed that coarse-to-fine-like responses are possible with feedforward mechanisms because a set of simulated STRF dynamics was able to produce responses resembling those appearing in V1 recordings. Indeed, manipulation of two parameters of the simulated STRFs, frequency and diameter, produced responses resembling most of the response varieties observed from recorded neurons. However, upon comparing simulated STRFs to ground-truth STRFs, several discrepancies appeared, suggesting that the model is not biologically plausible. Furthermore, modeling the response with STRFs that resembled ground-truth STRFs produced coarse-to-fine-like responses which were limited to the lower band of SFs. Ultimately, these results are inconclusive regarding whether coarse-to-fine processing is purely a feedforward or feedback process; however, the results do suggest that a feedforward process may facilitate coarse-to-fine processing while a feedback process may enhance responses to higher SFs.

# 3. Methods

## 3.1) Surgical procedure and running wheel acclimation

All experiments were conducted using awake C57BL mice (N(male) = 7, N(female) = 4) which were approved by the Animal Care and Use Committee of the University of Virginia. I followed laboratory surgical procedures previously published (Skyberg, R., Tanabe, S., Chen, H. & Cang, J., 2022). All procedures were performed in accordance with approved guidelines and regulations. A surgery was first conducted to implant the head plate. Anesthesia was induced via isoflurane (3-4% for induction, 2% for maintenance, in O2, ~0.5-0.6L/min, VetFlo, Kent Scientific). 0.1ml Atropine (0.3mg/kg in 10% saline) and 0.1ml dexamethasone (2mg/kg in 10% saline) were injected subcutaneously to reduce secretions and edema. Body temperature was continuously monitored and held at approximately 37 degrees Celsius via an external heating pad. Artificial tears (Henry Schein Medical) were applied to the eyes for protection during the surgery. After surgery, 0.2ml carprofen (5mg/kg) was injected subcutaneously for analgesia. The mice were kept in a heated chamber to recover from the surgery until ambulatory then transferred to their home cage and monitored daily for pain and wound health.

Following head plating, mice were acclimated to running on the wheel under head-fixed conditions for 4 days for 30-minute sessions per day. The first three days of running wheel acclimation were done with the absence of visual stimulation. On the fourth day, a sign switching (once per 500ms) checkerboard stimulus generated via the MATLAB Psychophysics toolbox was shown on an LCD monitor placed 25 cm away from the head fixation post (52.7cm x 29.6cm, 60Hz refresh rate, ~50cd/m2 mean luminance, gamma corrected) to acclimate mice to the presence of stimuli. On the day of the recording, a craniotomy was done under isoflurane anesthesia above the left primary visual cortex (~1.0 X 1.0 mm2, ~2.8mm lateral and ~0.5mm anterior from lambda). Recordings were done approximately 3 hours after craniotomy.

## 3.2) Visual stimulation

All stimuli were generated via the MATLAB Psychophysics toolbox and shown on the monitor setup previously mentioned. The monitor covered approximately a 128-degree azimuth and 72-degree elevation at the given distance from the head fixation post. The location of the receptive field was manually approximated via a flashing dot stimulus at the beginning of every recording. Once the receptive field was found, the monitor was carefully moved until stimuli were centered on the receptive field.

The main stimulus used for this experiment was an ensemble of sinusoidal gratings that changed at each frame at a refresh rate of 60hz. The stimulus was confined to a 50-degree diameter window in the center of the monitor. The gratings varied with respect to SF (0, 0.02, 0.04, 0.06, 0.08, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2, 0.22, 0.24, 0.26, 0.28, 0.3, 0.32), orientation (in degrees: 0, 30, 60, 90, 120, 150), and spatial phase (in radians: 0, $\frac{\pi}{2}$, $\pi$, $\frac{3\pi}{2}$) comprising 408 conditions (384 stimuli + 24 blanks). Gratings randomly transitioned from condition to condition with blank images of uniform luminance (0 SF) interleaved in between to provide a measure of the baseline neuronal firing rate. Each condition was repeated 200 times for a total recording time of ~24 min. This stimulus paradigm was inspired by the paradigm described by Bredfeldt and Ringach, 2002.

## 3.3) Physiological recordings and spike sorting

Recordings were done using high-density multielectron silicon microprobes developed by Sotiris Masmanidis from the University of California, Los Angeles (Yang et al., 2020). I used single probe 64M designs. The probe was carefully inserted into the craniotomy and lowered to a depth of ~0.8-0.9mm below the cortical surface. After reaching its final depth, the probe was left to settle for ~10-15 min before recordings began.

Voltage signals from the probes were band-pass filtered (300-6000 hz) and recorded at a sampling rate of 20kHz (RHD Evaluation System, Intan Technologies, RRID:SCR_017446). Spike waveforms were sorted offline using the software package MountainSort (Chung, et al., 2017; RRID:SCR_017446). Single spikes were identified using two clustering metrics given by the spike sorting algorithm. The first is noise overlap, which measures how much overlap the cluster has to a cluster composed of randomly sampled spikes from noise. The second is isolation, which indicates how well isolated a cluster is from other clusters. I classified single units as clusters whose noise overlap was <0.07 and isolation was >0.96. Visually responsive units were classified as units which showed a peak or valley in the trial averaged stimulus triggered response rate relative to a blank stimulus between 50-150ms time lag.

### 3.4) Stimulus triggered response functions

Grating stimuli varied with respect to three parameters: orientation ($\theta$), SF ($f$), and phase (ϕ). Each unique combination of the values of the three parameters produced a unique stimulus condition (s). Processed data from spike sorting was output as a point process of stimulus and spike onset times given in 20kHz time bins following the start of the recording. The onset times were binned into 1ms time bins by dividing the entire array by 20. Spike rasters centered on the onset time of each repeat of a stimulus condition were generated for each unit to produce a raw stimulus triggered response with respect to time, $R(s,t)_{raw}$. Each raster was thresholded to only include spikes occurring up to 200ms following the onset time of a stimulus. The vector sum of individual rasters corresponding to repeats of a stimulus condition was taken and divided by the number of repeats to produce a trial averaged stimulus triggered response $R(s,t)_{averaged}$. The same procedure was repeated for blank stimuli to produce an averaged blank triggered response $R(b,t)_{averaged}$. The vectors of $R(b,t)_{averaged}$ therefore served as a measure of the neuronal baseline firing rate. The vectors of $R(s,t)_{averaged}$ were marginalized with respect to orientation and phase by averaging over the two parameters to produce the raw stimulus triggered response functions with respect to time and SF, $R(f,t)_{raw}$. The vectors $R(f,t)_{raw}$ were then baseline subtracted ( $R(f,t)_{raw}$ - $R(b,t)_{averaged}$ ) to produce vectors describing the relative strength of the SF triggered response with respect to time $R(f,t)$. In essence, positive values of $R(f,t)$ at any time point indicate that a particular unit had a higher average firing rate to a stimulus of the given SF than to the blank stimulus. The opposite is true for negative values of $R(f,t)$. Finally, the vectors of $R(f,t)$ were smoothed by applying a 25ms rolling average.

### 3.5) Modeling of neuronal response profiles

Neuronal responses were simulated using a Linear-Nonlinear Poisson (LNP) cascade model. Given a spatiotemporal stimulus ensemble, X, and a simulated spatiotemporal receptive field (STRF), K, I first calculated the convolution between X and K:

$$Y(t) = \sum K(t - \tau_i)X(\tau_i)$$

The stimulus ensemble in this analysis is identical to the one used in the electrophysiological recordings. For this analysis, the convolution between X and K was calculated by creating a Toeplitz matrix of the stimulus ensemble, $X_{Toep}$ and taking the dot product of $X_{Toep}$ with K. Since X is a structure containing the grayscale values of each two-dimensional frame of the stimulus ensemble, it has a dimension of MxNxT, where M and N correspond to the height and width of each frame and T corresponds to the duration of the stimulus ensemble. Similarly, K is a structure containing the grayscale values of each frame of the simulated STRF and has a dimension of PxQxR. Each frame of the structures X and K was flattened to produce the 2D matrices $X_{flat}$ and $K_{flat}$ of dimension VxT and WxR respectively. Flattening was achieved by directly indexing the values falling within the 50-degree circular window of the X and K. This method of flattening did not affect the final calculation. $X_{Toep}$ was therefore generated using $X_{flat}$ and thus had a dimension of V*RxT. The convolution between the X and K was then ultimately calculated as:

$$Y(t) = X_{Toep} \cdot K_{flat}$$

The output of the convolution is an array of values with a length equal to T. Because the convolution produces negative values, it is passed through an exponential function to generate a function of the response rate with respect to time. This function is additionally scaled by a constant factor of 0.0125 to constrain the range between extreme high and low values that result from the exponential operation:

$$R(t) = 0.0125e^{Y(t)}$$

The simulated STRFs, used to generate the response rate functions were a series of two-dimensional Gabor filters given by the following equation:

$$g(x, y, \theta, f, \phi, \sigma, \gamma) = e^{\frac{x'^2 + \gamma y'^2}{2\sigma^2}} \cos(2\pi x' f + \phi)$$

Where x' and y' are transformations of the coordinate space applying a rotation of the given orientation $\theta$:

$$x' = x\cos(\theta) + y\sin(\theta)$$

$$y' = -x\sin(\theta) + y\cos(\theta)$$

$\sigma$ is the standard deviation of the Gaussian envelope, $\gamma$ is the aspect ratio of the Gabor, and $f$ and $\phi$ are the frequency and the phase of the sinusoidal component respectively. The filters used for all the modeling conditions had constant $\theta$ values of 0-degrees and constant $\phi$ values of -90-degrees. Because the stimulus is discretized into 60hz bins, the duration of the simulated STRFs was chosen to be 13 frames long to approximate the 200ms length of the stimulus triggered response functions. The first 7 frames of the simulated STRFs were set as blanks to simulate the natural lag in the neuronal response. The last 6 frames of the simulated STRFs contained Gabor filters which changed from frame to frame with respect to the values of the $f$

and σ parameters. The way the parameters changed with over time formed curves which took on one of three shapes: a power law shape, logarithmic shape, or a linear shape (Fig 1). The values for these curves were calculated by passing power law, logarithmic and linear functions through a minmax normalization function with specifically chosen upper and lower bounds:

$$minmax(f(x), L, U) = L + \frac{f(x) - min(f(x))}{max(f(x)) - min(f(x))}(U - L)$$

Where L denotes the lower bound and U denotes the upper bound of the function. The following equations describe each of the individual curves:

$$f_{powerlaw} = minmax\left(-\frac{1}{x}, 0.04, 0.2\right)$$

$$\sigma_{powerlaw} = minmax\left(\frac{1}{x}, 1, 5\right)$$

$$f_{log} = minmax(log(x), 0.04, 0.2)$$

$$\sigma_{log} = minmax(-log(x), 1, 5)$$

$$f_{linear} = minmax(x, 0.04, 0.2)$$

$$\sigma_{linear} = minmax(-x, 1, 5)$$

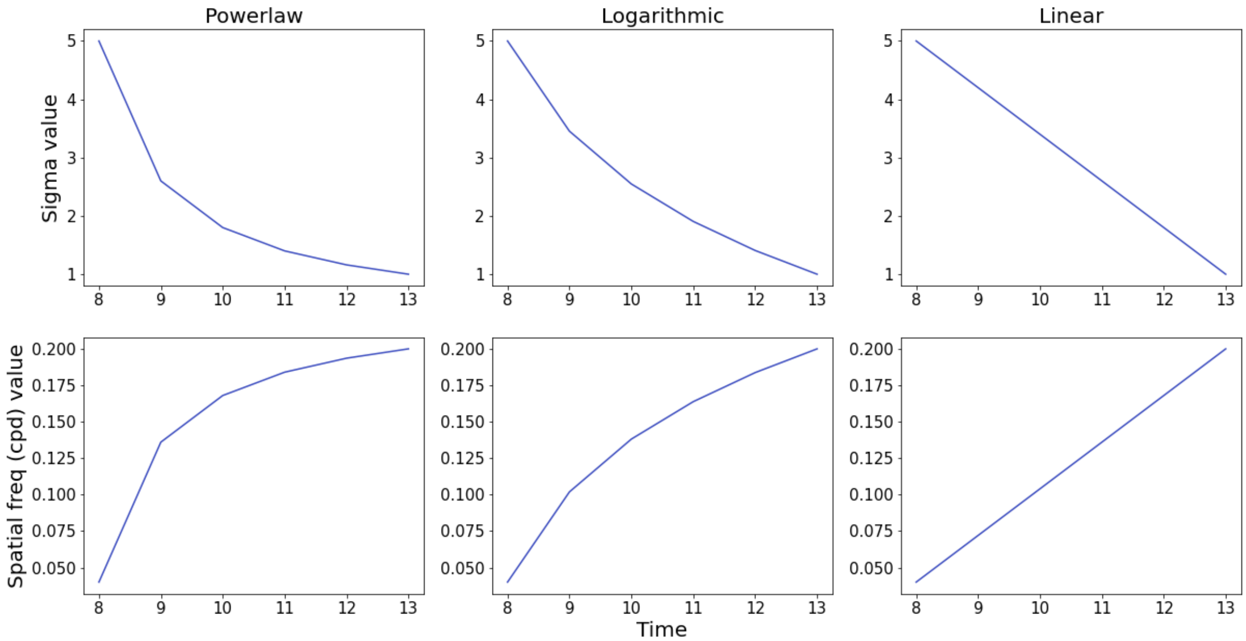


*Figure 1: A schematic illustrating how the Gabor filters used in the STRFs change over time with respect to σ (The standard deviation of the Gaussian window) and $f$ (The frequency of the sinusoidal component). For condition 1, σ and $f$ co-evolved in every possible combination of curves shown in the figure (see Table 1).*

The way that $f$ and σ co-evolved comprised three conditions. In the first condition, $f$ and σ co-evolved in every one of 9 possible combinations of the three curves (Table 1). In the second condition, σ was held at a constant value of 3 while $f$ changed according to the three curves (Fig 2). Conversely, in the third condition $f$ was held at a constant value of 0.06 and σ changed according to the 3 curves (Fig 3). The range of σ values was chosen because it closely approximated the range of receptive field sizes mapped from recorded units. Similarly, the range of $f$ values was chosen because it closely approximated the range of SFs to which recorded units were responsive.

| | $f_{powerlaw}$ | $f_{log}$ | $f_{linear}$ |
|---|---|---|---|
| $\sigma_{powerlaw}$ | $f_{powerlaw}\ x\ \sigma_{powerlaw}$ | $f_{log}\ x\ \sigma_{powerlaw}$ | $f_{linear}\ x\ \sigma_{powerlaw}$ |
| $\sigma_{log}$ | $f_{powerlaw}\ x\ \sigma_{log}$ | $f_{log}\ x\ \sigma_{log}$ | $f_{linear}\ x\ \sigma_{log}$ |
| $\sigma_{linear}$ | $f_{powerlaw}\ x\ \sigma_{linear}$ | $f_{log}\ x\ \sigma_{linear}$ | $f_{linear}\ x\ \sigma_{linear}$ |

*Table 1: The first row shows functions of the $f$ parameter. The first column shows functions of the σ parameter. Individual cells show the possible combinations between $f$ parameter functions and σ parameter functions.*

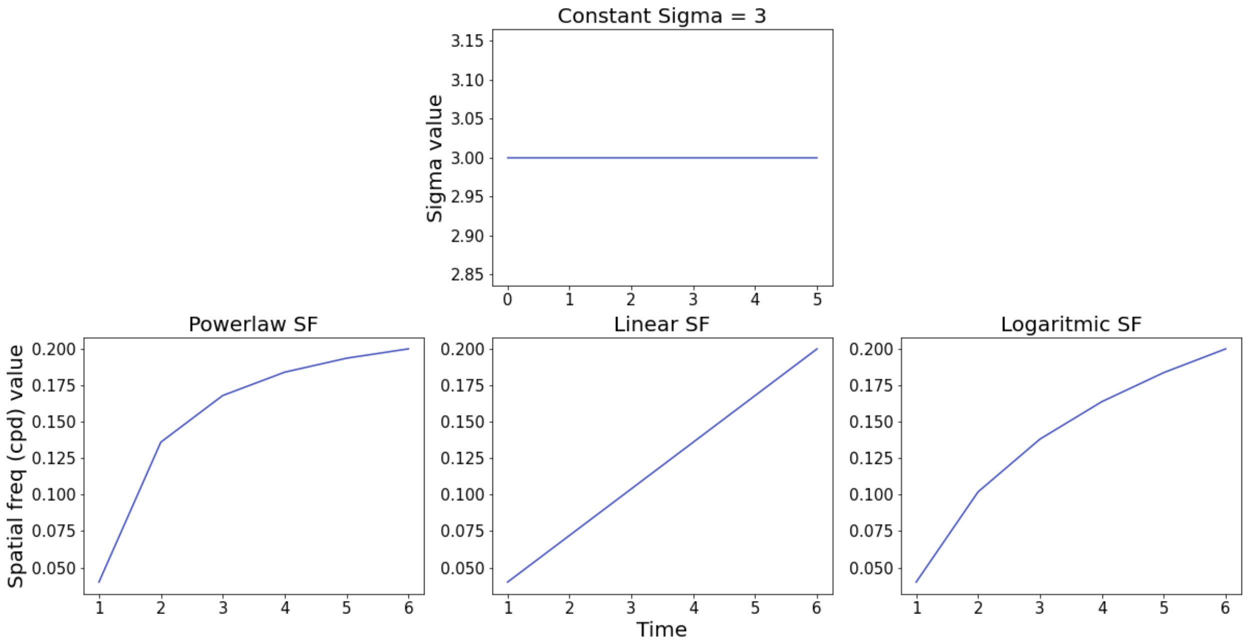


*Figure 2: A schematic illustrating how the values of the $\sigma$ and $f$ parameters change over time in condition 2. $\sigma$ remains at a constant value of 3 while $f$ could take on a power law, logarithmic or linear shape.*

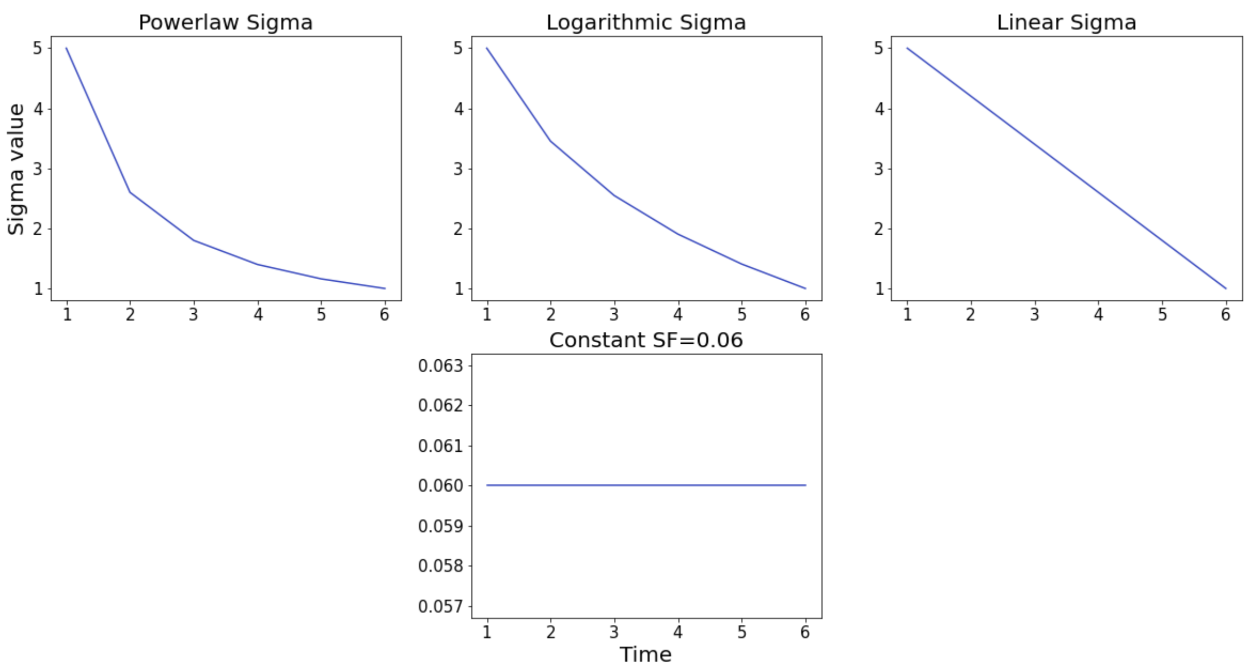


*Figure 3: A schematic illustrating how the values of the $\sigma$ and $f$ parameters change over time in condition 3. $f$ remains at a constant value of 0.06 while $\sigma$ could take on a powerlaw, logarithmic or linear shape.*

After generating the rate functions with the simulated STRFs described above, the rate functions were used to generate vectors, $R(f,t)_{sim}$, in an equivalent fashion to how they were generated in section 3.4.

## 3.6) STRF mapping

For each unit, a trial averaged stimulus triggered response function is calculated prior to marginalization ( $R(s,t)_{averaged}$ ) for all 384 non-blank stimulus conditions over a duration of 200ms. The response functions are subsequently baseline subtracted ($R(s,t)_{averaged} - R(b,t)_{averaged}$ ) to produce vectors describing the relative strength of the stimulus triggered response with respect to time and each unique stimulus condition, $R(s,t)$. Furthermore, each baseline subtracted response function is smoothed with a 20ms rolling average to reduce the impact of noise. Finally, for each unit, a matrix $W$ (Fig 4) is generated using the function $R(s,t)$ such that:

$$W = \begin{bmatrix} R(s=1,t=1) & \cdots & R(s=1,t=200) \\ \vdots & \ddots & \vdots \\ R(s=384,t=1) & \cdots & R(s=384,t=200) \end{bmatrix}$$

In essence, each row of W is a vector corresponding to the unit's response to a unique stimulus condition. Furthermore, a three-dimensional matrix C (Fig 4) is generated to describe the spatial features of each stimulus condition. Matrix C is composed of a series of two-dimensional matrices each containing the grayscale values of a sinusoidal grating corresponding to each stimulus condition. Matrix C has a dimension of MxNxS, where M and N correspond to the height and width of each stimulus frame and S corresponds to the stimulus condition id and has a length of 384. To produce the STRF for a particular unit, I simply take the dot-product of matrix C and the matrix W generated for the unit. I.e.:

$$STRF_{unit} = C \cdot W_{unit}$$

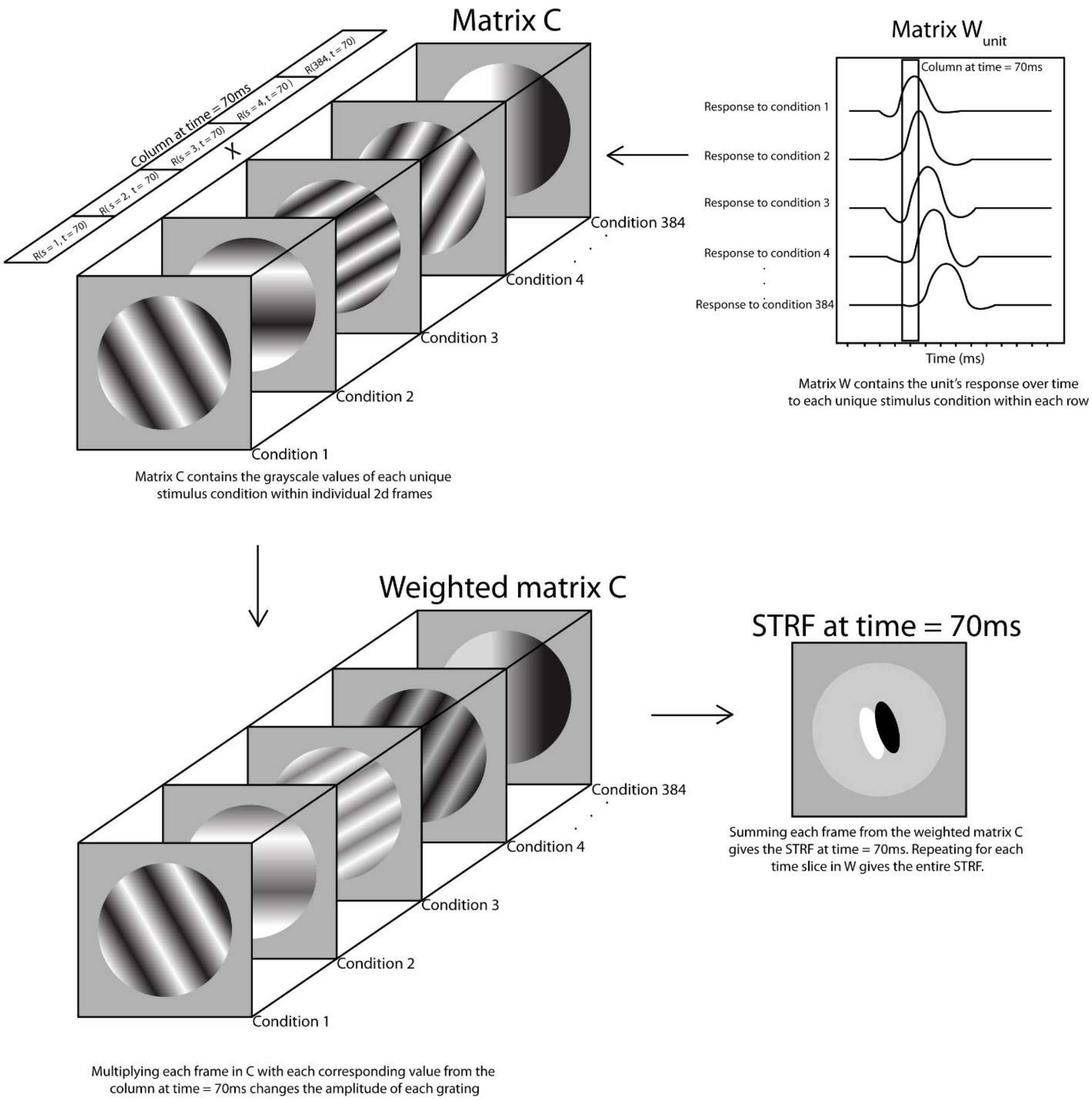


*Figure 4: Schematic demonstrating how the STRFs are generated for each unit. Matrix C is a structure containing the grayscale values of gratings corresponding to each stimulus condition. Matrix W is a structure containing the unit's response over time to each condition. The dot product of C and W produces the STRF, a structure containing a picture of the spatial receptive field at each timepoint of the response.*

Essentially, this operation is a linear combination of C with each column of W and produces a new three-dimensional matrix $STRF_{unit}$. Put another way, for each value in a column of W, I take the product of the value with every corresponding frame in C and then sum the frames to create a new frame (fig 4). I then concatenate the new frame to $STRF_{unit}$ and subsequently repeat this process for each column in W. $STRF_{unit}$ is composed of a series of two-dimensional matrices each containing the grayscale values of the unit's spatial receptive field at each timepoint of the unit's response. In essence, $STRF_{unit}$ has a dimension of MxNx$t_{max}$ where M and N correspond to the height and width of each STRF frame and $t_{max}$ corresponds to the duration of the STRF and has a length of 200.

### 3.7) Quantification of STRF parameters

To quantify the parameters of STRFs mapped from recorded units, Gabor filters were fitted to each frame of a unit's STRF. Fitting of Gabor filters was achieved using cross-validated mean squared error (MSE) minimization. For each frame of a unit's STRF a Gabor filter was generated using an initial guess of parameters bounded within an expected range. The STRF frame and the Gabor filter were then subsequently flattened. The flattened vectors were subtracted and squared, and the mean of the resulting vector was calculated to produce a MSE score. The MSE score was then minimized via the minimize function from the Scipy optimize Python library using the Nelder-Mead algorithm (Singer, S., & Nelder, J., 2009) to fit the parameters of the Gabor filter to the STRF frame. Fitting was repeated 30 times and the Pearson's correlation score between each fit and the STRF frame was calculated and stored for cross-validation. The parameters from the fitting which produced the highest Pearson's correlation score were retrieved and used as the best fitting parameters.

Because the neuronal response to a stimulus occurs at a delay and is transient, the STRF is only visible for a short duration which was labeled manually and defined as the active range. Active ranges varied from unit to unit and typically spanned 70ms to 130ms. Only the parameters retrieved from STRF frames falling within the active range were used for analysis in this study. This analysis focuses on the $\sigma$ and $f$ parameters recovered from the fits. Functions of how these parameters change with respect to time, $f_{\text{fit}}$ and $\sigma_{fit}$, were generated to compare the spatiotemporal dynamics of mapped STRFs to the simulated STRFs. Because $f_{\text{fit}}$ and $\sigma_{fit}$ had some noise, they were smoothed using a second order Savitzky-Golay filter with a window equivalent to the length of each active range (the window was subtracted by 1 if the length of the active range was even). Savitzky-Golay filtering was applied using the savgol_filter function from the Scipy signal Python library. Savitzky-Golay filtering was applied only for visualization and only the original values were used for data analysis.

# 4. Results

## 4.1) Stimulus triggered response profiles

Stimulus triggered response functions with respect to SF and time $R(f,t)$ were generated for 72 units total by taking 200ms of the unit's response following stimulus presentations, averaging across repeats of the same stimulus condition, and marginalizing across orientation and phase. The criterion for visually evoked responses (VERs) was defined as responses which rose above the mean response rate during the first 40ms + 4 standard deviations of the mean response rate during the first 40ms. This VER criterion is identical to the one used in Bredfeldt, C. E. & Ringach, D. L. 2002. For each unit, onset time of a response to a particular SF was calculated as the time at which the response rate first rose above the VER criterion. Across the population of units analyzed in this study, onsets to higher SFs occurred later than onsets to lower SFs (fig 5a & 5b). This response phenomenon agrees with the results shown in (Skyberg, R., Tanabe, S., Chen, H. & Cang, J., 2022) and is defined as the coarse-to-fine response. These results also show that the temporal progression of responses to SFs occurs in a smooth and continuous fashion rather than a discrete and abrupt fashion (fig 5b & 5c). For each unit, linear regressions were fit to the distributions of onset times with respect to SF (fig 5c). The mean of the coefficients retrieved from the linear regressions showed that, on average, neurons in V1 shift their preferred SF positively by 0.004cpd per millisecond (fig 5b).

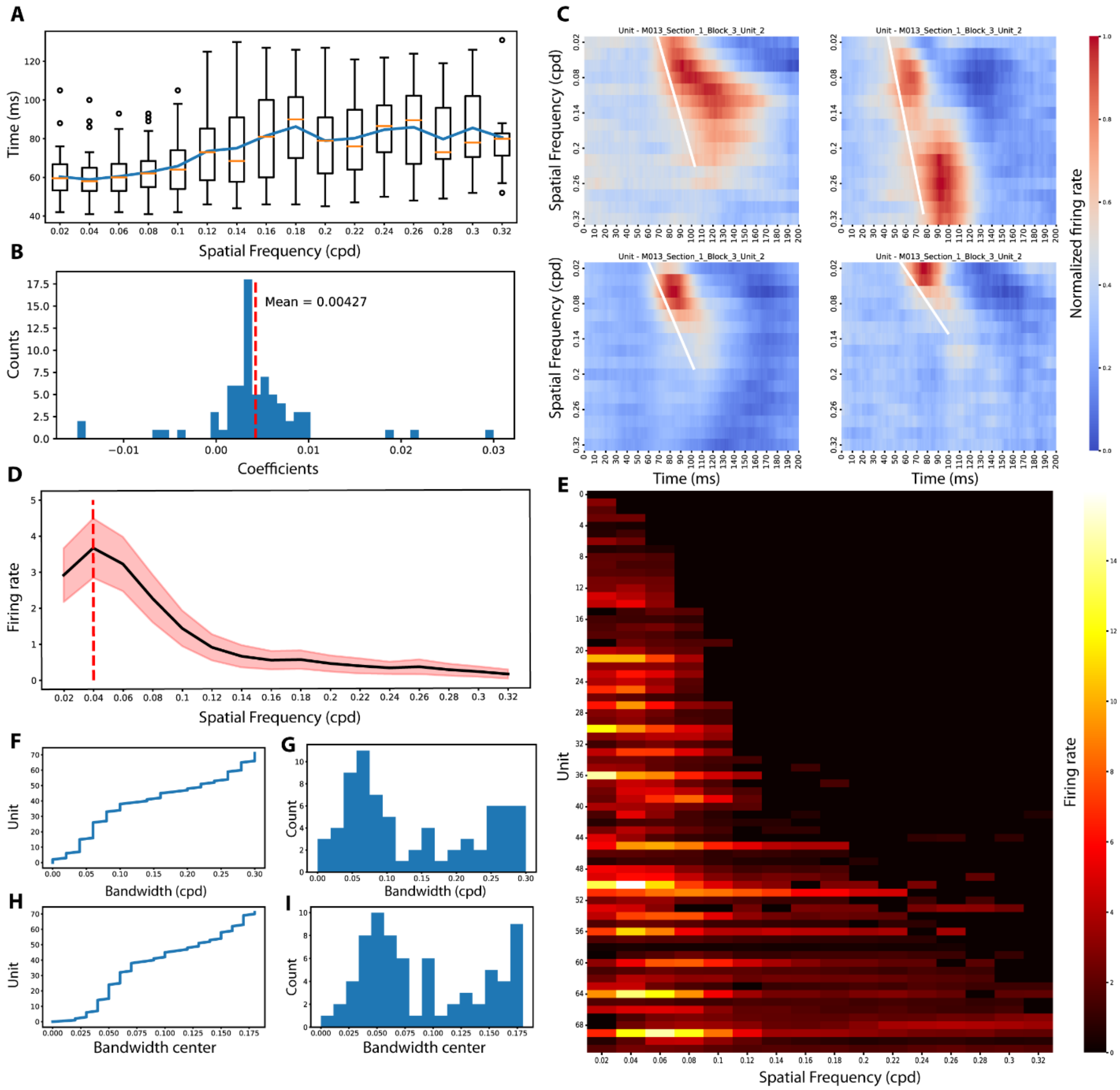


*Figure 5:* ***Analysis of coarse-to-fine responses*****.** **A:** Distribution of onset times with respect to SF for the population of units observed in the study. The blue trace indicates the mean onset time for each SF. Onset times for higher SFs occur later. **B:** Histogram of coefficients from linear regression fits to onset times. **C:** Example $R(f, t)$ responses from four units included in the study. The white line is the linear regression fit to onset times with respect to spatial frequency. The length of each white line indicates the bandwidth of the unit. **D:** Population averaged spatial frequency tuning curve. **E:** Individual unit spatial frequency

tuning curves calculated with the method described in the text, sorted in ascending order from lowest bandwidth to highest bandwidth. **F:** Bandwidth of each unit sorted in ascending order. **G:** Histogram of bandwidths. **H:** Bandwidth center of each unit sorted in ascending order. **I:** Histogram of bandwidth centers.

Coarse-to-fine responses tended to have an enhanced response to low SFs (0.02 – 0.16 cpd) compared to high SFs (0.18 – 0.32 cpd) (fig 5d). To quantify how the response rates differed between low SFs and high SFs, I defined the response window as all the time points in which response rates rose above the VER criterion. Subsequently, I calculated the average response rate to each SF over the entire response window to create SF tuning curves for each unit (fig 5e). If no responses rose above the VER criterion for a particular SF, the average response to that SF was set as 0. Among all the units, 46 only responded to low SFs and 26 responded to both low and high SFs (fig 5e). Among the units which only responded to low SFs, 12 showed no temporal SF dynamics. Among the units which responded to low and high SFs, there was a wide variety in the ratio of peak average response rates between low and high SFs. A total of two units had slightly higher peaks to high SFs. A total of 8 units had low SF peaks 1-2x higher than high SF peaks. Another set of 8 units had low SF peaks 2-3x higher than high SF peaks. 5 units had low SF peaks 3-5x higher than high SF peaks and 3 units had low SF peaks 5-10x higher than high SF peaks. Taking the average of all tuning curves showed that the population is tuned to 0.04cpd on average (fig 5d).

For each unit, the bandwidth was calculated as the difference between the highest SF to elicit a response above the VER criterion and the lowest SF to elicit a response above the VER criterion. The distribution of bandwidths from the population of units analyzed in this study spanned the entire range of possible bandwidths (fig 5e & 5f). Furthermore, a histogram of the bandwidths shows that there is a bimodal distribution of bandwidths in the population of units analyzed in this study (fig 5g). Additionally, the center of the bandwidth for each unit was defined as the midpoint between the lowest and highest SF to elicit a response above the VER criterion. The distribution of bandwidth centers was nearly identical to the distribution of bandwidths, indicating that as the bandwidth increased, the center shifted towards higher SFs. Furthermore, across the whole variety of response profiles that were observed, over half of the units had responses to low SFs that were suppressed near onset times which correlated with high SFs. This phenomenon of late suppression of low SFs agrees with observations from multiple studies. (Bredfeldt, C. E., & Ringach, D. L., 2002; Purushothaman, G., Chen, X., Yampolsky, D. & Casagrande, V. A., 2014; Skyberg, R., Tanabe, S., Chen, H. & Cang, J., 2022). The following experiments are an attempt to investigate the mechanisms behind the variety of response profiles observed.

### 4.2) Modeling of the neuronal response profiles

For this analysis, I assume that coarse-to-fine processing is the result of classical hierarchical feedforward processing described by Hubel, D. & Wiesel, T., 1962. Given this assumption, I hypothesized that the coarse-to-fine response could be observed in the STRFs of V1 neurons as dynamic changes occurring in spatial structure over short timescales. To explore this possibility, I first wanted to explore what STRF dynamics are necessary to produce response profiles resembling those observed from recorded V1 neurons. I therefore modeled responses using a Linear-Nonlinear Poisson (LNP) cascade model. This type of model was chosen for its simplicity and because it can be used to predict how biological neurons with specific receptive fields respond to a variety of stimuli. It should be noted that because the LNP model is a type of generalized linear model, it is unable to recreate nonlinear response dynamics present in real neurons. This analysis is also limited by the fact that only simple-cell receptive field structures were used. Nonetheless, this analysis is informative because it can test whether coarse-to-fine processing is caused by feedforward mechanisms. Essentially, if the simulated STRFs used in the model produce coarse-to-fine responses and are biologically plausible, this would constitute strong evidence that coarse-to-fine processing occurs via feedforward mechanisms.

Feedforward mechanism dependent changes in STRF structure have previously been observed in the V1 of cats (Suder, K., Funke, K., Zhao, Y., Kerscher, N., Wennekers, T. & Wörgötter, F., 2002). Over stimulus presentations of 300ms in duration, the diameters of STRFs in V1 neurons shrank by 22% on average. These reductions in STRF diameter were caused by a transition of dLGN firing from phasic to tonic, implicating feedforward activity as a causal mechanism of the observed STRF changes. Given this finding and the fact that the coarse-to-fine response is a SF dependent phenomenon, I designed the simulated STRFs as series of 2D Gabor filters which varied over time with regards to the $\sigma$ parameter, governing the diameter of the filter's Gaussian envelope, and the $f$ parameter, governing the frequency of the filter's sinusoidal component. The Gabor filters appeared within the last 6 frames of a series of 1-frames; the first seven frames of this series were blank (fig 6a). This temporal design was chosen for two reasons. The first reason is that the stimulus ensemble used in experiments had a refresh rate of 60hz and the stimulus triggered responses were 200ms long. A 13-frame long simulated STRF approximated the 200ms duration of the stimulus triggered responses binned at 60hz. Second, responses to stimuli in real neurons typically appeared within a range between 65 - 135ms. Designing the simulated STRFs such that the active range spanned the last 6 frames allowed the model to generate responses with similar ranges (fig 6a & 6b).

The way that the values f and σ changed over the course of the simulated STRF active range was generated by either a power law function, logarithmic function, or linear function with respect to time (Fig 1). For f values, the trend was increasing and for σ values the trend was decreasing. For all of the generating functions, the range of $f$ values was bounded between 0.04 and 0.2 and the range of σ values was bounded between 1 and 5. These values were chosen because they approximately matched values seen in STRFs mapped from recorded neurons. The temporal dynamics of simulated STRF parameters fell into one of 3 conditions (Fig 1). In condition 1, the values $f$ and σ both varied over the duration of the active range and could take on any of the 9 combinations of generating functions (Table 1). In condition 2 (Fig 2), $f$ values were produced by any of its 3 generating functions and σ remained at a constant value of 3. Condition 3 (Fig 3) was the opposite of condition 2; $f$ remained at a constant value of 0.06 and σ values were produced by any of its 3 generating functions.

In Condition 1, where σ decreased as $f$ increased, the model produced response profiles in which the peak response to low SFs was ~3x higher than the peak response to high SFs (fig 6a & fig 7a). The response rate to high SFs and the band of high SFs to which a response occurred depended on two factors. The first factor was the rate of change of $f$ values. In cases where $f$ values changed faster than σ values, the model generated higher responses to high SFs and responses occurred to wider band of SFs. In contrast, if $f$ values changed more slowly than σ values, the model generated lower responses to high SFs and responses occurred to narrower band of SFs (fig 7d & 7e). The second factor was the contribution of the rate of change of σ values, which had an antagonistic effect on the $f$ parameter contribution. Generally, if σ values had a faster rate of change, the model's response to high SFs was suppressed (fig 7d & 7e). For example, in the extreme case where $f$ increased linearly and σ decreased in a power law fashion ($f_{linear}$ x $\sigma_{powerlaw}$), the simulated responses resembled coarse-to-fine response profiles that were limited to the lower band of SFs (fig 6b). The results from this condition suggest that STRF diameter and frequency have an antagonistic relationship such that responses to high SFs only occur when frequency increases at a faster rate than diameter decreases.

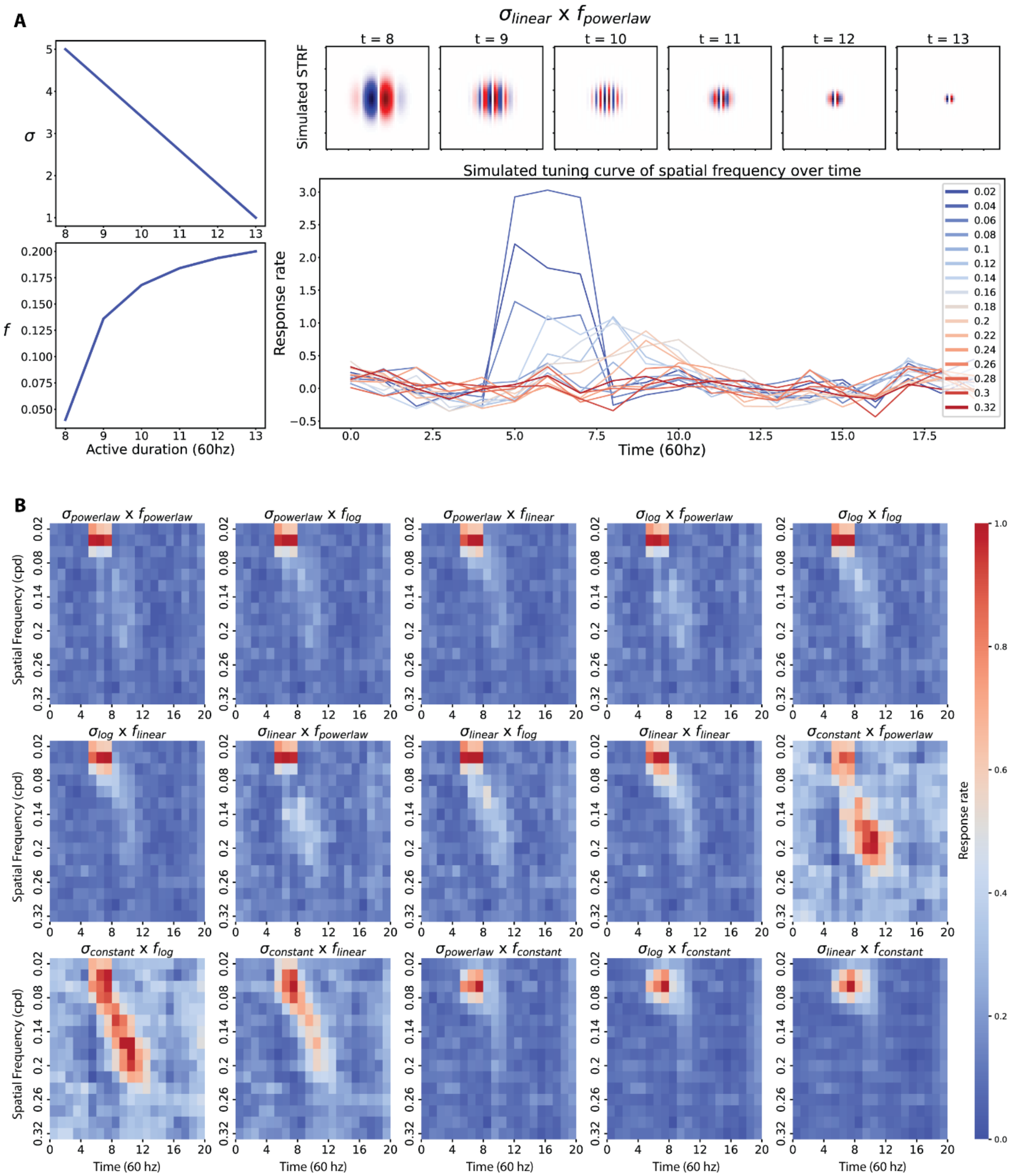


*Figure 6:* ***Model responses.*** **A:** An example response showing how the $\sigma$ and $f$ parameters change over the active duration (left top and bottom panels, a visual representation of the simulated STRF (top 6 panels on the right), and the $R(f, t)$ response from the modeling condition. **B:** $R(f, t)$ responses from all modeling conditions explored in the study.

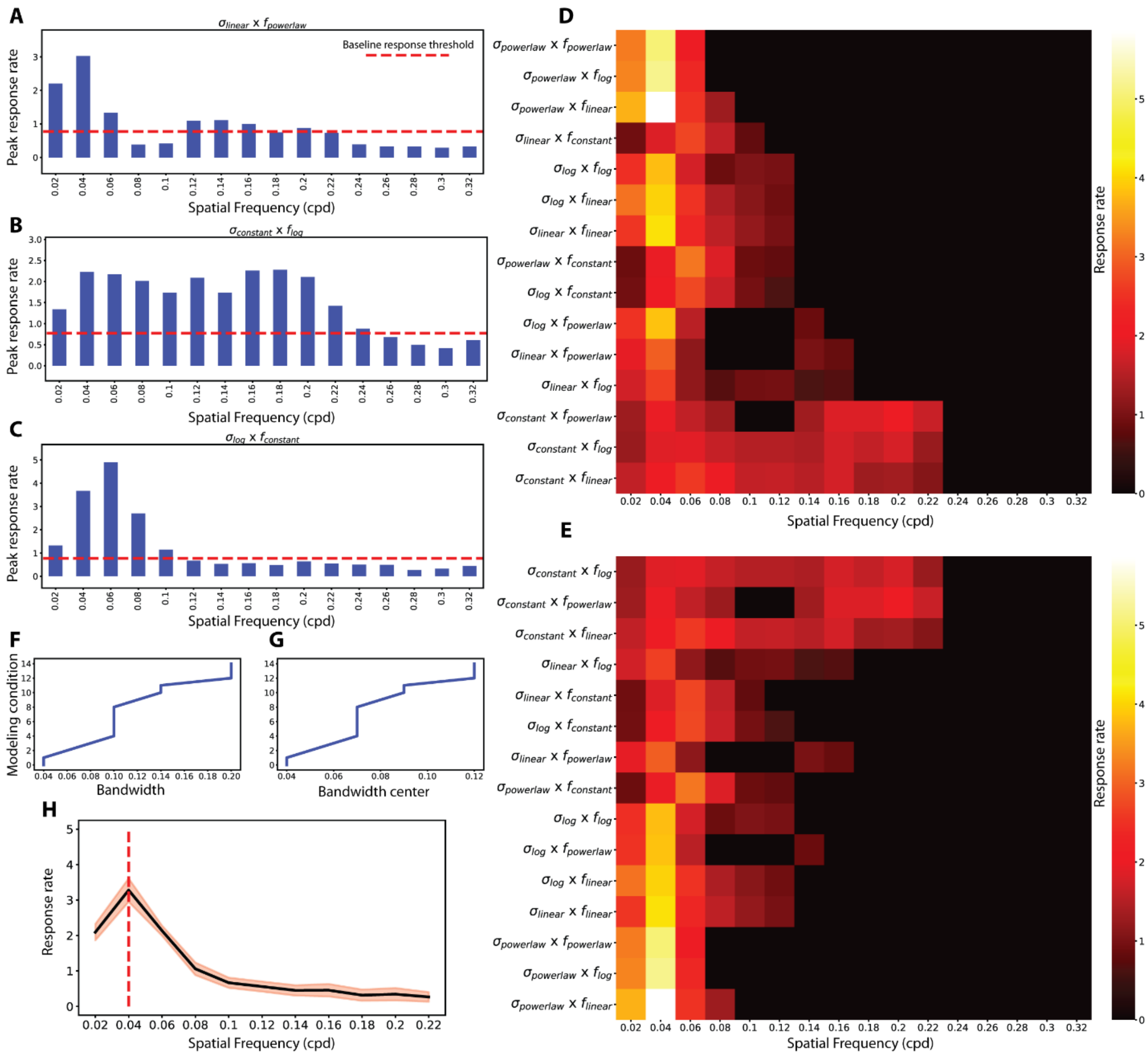

*Figure 7: **Analysis of the model responses.*** **A:** Peak response rate at each spatial frequency taken from an example from condition 1. Red line indicates the baseline response threshold defined in an identical manner to the VER criterion from section 4.1. **B:** Same as **A** but from an example response from condition 2. **C:** Same as **A** but from an example response from condition 3. **D:** Individual response spatial frequency tuning curves calculated in an identical manner to the one in section 4.1, sorted in ascending order from lowest bandwidth to highest bandwidth. **E:** Same as **D** sorted in ascending order from lowest peak response rate to highest peak response rate. **F:** Bandwidths of responses sorted in ascending order. **G:** Bandwidth centers of the responses sorted in ascending order. **H:** Spatial frequency tuning curve averaged over all responses.

In condition 2, the model produced low and high SF responses with nearly equivalent peaks (fig 7b). Because σ remained constant in this condition, it had no antagonistic effect on the contribution of the $f$ parameter. If $f$ values had a high rate of change, the model generated response to high SFs response that were slightly higher than the responses to low SFs. If $f$ values had a low rate of change, the model generated responses to high SFs that were slightly lower than the response to low SFs. If $f$ values had a stable rate of change, the model generated response to high and low SFs that were essentially equivalent (fig 7d & 7e). The results from this condition suggest STRF diameter must remain constant for the high SF peak to be equal to or higher than the low SF peak.

In condition 3, the model produced response profiles in which the response only occurred to low SFs (fig 7c). Furthermore, response onset times to each low SF were relatively similar (fig 6b). Although this condition did not elicit any temporal SF dynamics, it did influence the width of responses. When $\sigma$ had a slower rate of change, responses became wider and vice versa. The results from this model condition suggest that STRF frequency must remain constant to produce responses without temporal SF dynamics.

The bandwidths of the responses generated by the model were analyzed in an identical fashion to section 4.1. Bandwidths had a similar distribution to the population of units analyzed in the study but were limited to a range between 0 and 0.2cpd (fig 7d, 7f & 7g). The modeled responses also tended to have enhanced responses to low SFs compared to high SFs (fig 7h). Overall, the results from the modeling indicate that short timescale STRF changes can produce coarse-to-fine-like responses. The responses generated from the model closely resemble a large subset of the response profiles observed from extracellularly recorded neurons. The model was, however, unable to reproduce fully suppressive response profiles and was unable to reproduce the low SF suppression observed during the high SF response. Nonetheless, these results provide support for the idea that coarse-to-fine responses may arise from feedforward processing. In the next section, I explore whether STRF structures from extracellularly recorded neurons resemble the simulated STRF structures which produced coarse-to-fine responses in the model.

### 4.3) Quantification of STRF parameters

If coarse-to-fine responses occur because of feedforward mechanisms, at least three assumptions must be true. The first assumption is that there must exist some STRF dynamics which can recreate the coarse-to-fine response profiles observed from recorded units. The second assumption is that these proposed STRF dynamics must be biologically plausible. The third assumption is, for units which exhibited coarse-to-fine response profiles, coarse-to-fine STRF dynamics must also be observed. The LNP model has shown that the evolution of the $f$ and $\sigma$ parameter values of a set of Gabor filters over time can recreate a large repertoire of the response profiles observed from recorded units. I next set out to test whether the latter two assumptions hold true. To do this, I first mapped STRFs from recorded units by transforming the neural responses from the frequency domain to the spatial domain. I subsequently fit Gabor filters to the mapped STRFs. Finally, I extracted the $f$ and $\sigma$ parameter values from the fits to quantify how the mapped STRFs change over time in reference to the parameters used in the model.

Mapping of STRFs was achieved by transforming the neuronal firing signal from the frequency domain to the spatial domain (ie: an inverse Fourier transform). The first step to achieve this was to generate a 3-dimensional matrix, C. This matrix contained 2-dimensional frames corresponding the grayscale values of each unique grating condition used for visual stimulation. Furthermore, a 200ms stimulus triggered response to each of these unique conditions was calculated; each of these responses formed the rows of a 2-dimensional response matrix, W. The second step was to calculate the linear combination of the frames of C with each column of W. Each of these linear combinations formed a frame representing the STRF at a single timepoint. This calculation is essentially equivalent to taking the dot product of C and W (fig 4). This method of mapping the STRFs was successful only in neurons with sufficiently linear responses. Clear STRFs were successfully mapped in ~60/72 neurons.

Furthermore, the following analysis was restricted to simple-cells because complex-cell STRFs do not have Gabor-like shapes. This further restricted the analysis down to 22 of the recorded units. This subset of units had STRFs with clear parallel, flanking excitatory and suppressive subregions which followed a Gabor-like pattern (fig 8b & 8f). The STRFs of these units typically appeared within a range between 65 – 90ms and disappeared within a range of between 100 – 135ms (mean onset time = 78.64±1.25ms; mean offset time = 115.91±2.22ms). The STRFs were active for a mean duration of 37.27±1.63ms (fig 8a).

To quantify how these STRFs changed over time, Gabor filters were fitted to each timeframe of the STRFs using mean-squared-error (MSE) minimization. For each frame in time of a STRF, a Gabor filter was initialized with a random guess of parameters bounded within an expected range. The MSE between the STRF frame and the Gabor filter was then minimized to produce a Gabor filter that best fit the frame. A 30-fold cross-validation was then performed by calculating the Pearson correlation between the Gabor fit and the STRF frame. The parameters of the fits with the highest cross-validation score ($CV_{max}$) were used for quantifying how the STRFs changed over time (fig 8b). The fitting procedure produced a function of $CV_{max}$ with respect to time for each unit analyzed, $CV(t)$. The $CV(t)$ functions for all 22 units produced a mean curve which peaked at 90ms with a value of 0.62±0.02 (fig 8c). Therefore, at 90ms fitting accuracy peaked because STRFs had less noise and became more Gabor-like on average.

I next wanted to test whether parameter values from the fits differed significantly between the onset time and offset time of the STRF active ranges. Active ranges for each STRF were defined manually and Gabor filters were subsequently fit to only the STRF frames of the active range. The σ and $f$ parameters from the best fits to each frame of the active range were retrieved to produce functions describing how the parameters change over the duration of the active range. These functions are respectively labeled as $\sigma_{fit}$ and $f_{\mathrm{fit}}$ (fig 8f, left top and bottom panels). The values of these functions differed significantly between onset and offset. Typically,

$f_{\mathrm{fit}}$ became higher over time, with an average onset value of 0.03±0.1 and an average offset value of 0.06±0.2 (paired t-test, p = 0.0002) (fig 8d). In contrast, $\sigma_{fit}$ typically became smaller over time with an average onset value of 3.34± 0.71 and an average offset value of 2.2±0.52 (paired t-test, p = 0.0003) (fig 8e). Therefore, STRFs in analyzed units generally increased with respect to $f$ and decreased with respect to σ, suggesting that coarse-to-fine dynamics are indeed observable in the STRF. However, this general pattern was mostly inconsistent with the modeling paradigm which produced coarse-to-fine-like responses because the range of $f$ and σ values used in the model was much wider. Furthermore, $f_{\mathrm{fit}}$ and $\sigma_{fit}$ typically changed at similar rates, yet responses to high SFs were still observed (fig 6b).

In a few instances, the values of $f_{\mathrm{fit}}$ and $\sigma_{fit}$ seemed to follow trends that recapitulated some of the paradigms used in the LNP model. For instance, units which only responded to low SFs with no difference in response delay had $f_{\mathrm{fit}}$ curves which mostly remained stable over time as expected from the model. Furthermore, in a representative unit where the low SF response peak was 2x higher than the high SF response peak, $f_{\mathrm{fit}}$ increased from 0.04 to 0.16 and $\sigma_{fit}(t)$ decreased from 4 to 1 in a linear-like fashion over time (fig 8f). Based on the results from the model, this range of $f_{\mathrm{fit}}$ increase is expected, however the range of $\sigma_{fit}$ decrease is not. I tested whether the LNP model can produce a similar response to this representative unit with similar parameter values and observed a coarse-to-fine-like response which was bounded to the lower range of SFs (fig 8g). Therefore, the simulated STRFs which were able to reproduce responses akin to the representative unit do not appear to be biologically plausible.

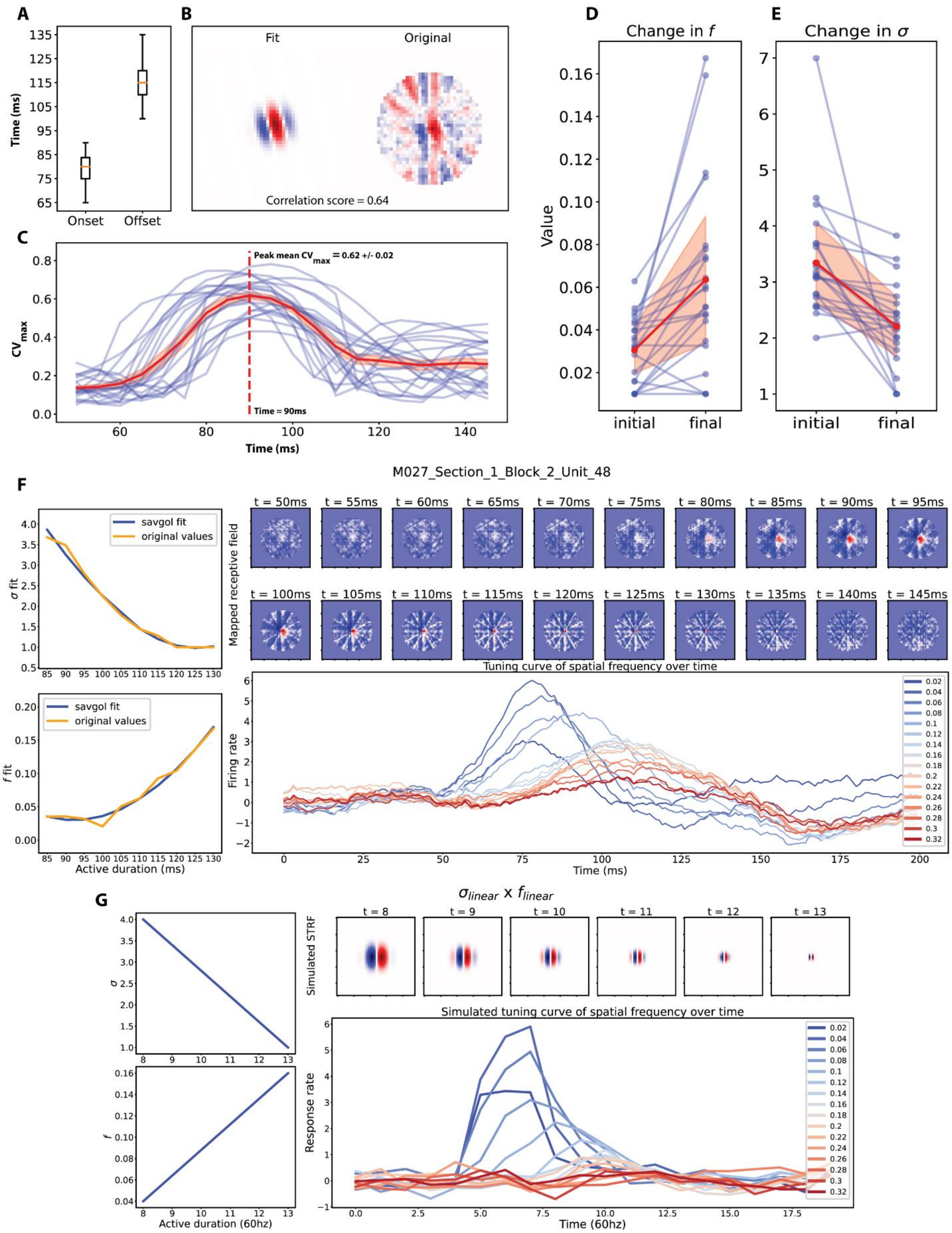


*Figure 8:* ***Quantification of STRF parameters.*** **A:** Distribution of onset and offset times for STRFs. **B:** Example of a Gabor fit to a mapped STRF using the method described in the text. **C:** $CV_{max}$ curves for all fitted STRFs. **D:** Difference in initial and final $f$

values retrieved from the Gabor fits. **E:** Difference in initial and final $\sigma$ values retrieved from the Gabor fits. **F:** Data from an example neuron. Left top and bottom panels show the $\sigma_{fit}$ and $f_{fit}$ curves from the Gabor fitting procedure. The series of panels in the upper right show the mapped STRF for the unit. The bottom panel shows the R($f$, t) response curves for each SF. **G:** Data from an attempt to model the response in **F.** Left top and bottom panels show how the $\sigma$ and $f$ parameters change over the active duration, the series of panels in the top right show a visual representation of the simulated STRF, the bottom panel shows and R($f, t$) responses from the modeling attempt. The modeling attempt was only able to capture the low SF portion of the response.

Due to the limited amount of data, most units with a high SF response that were analyzed had low response rates. This made it difficult to assess whether there was a relationship between $f_{\mathrm{fit}}$ and $\sigma_{fit}$ values and high SF response rate. However, I had some expectations of what the relationship should look like based on the results of the model. If high SF response rate is function of the $f_{\mathrm{fit}}$ rate of change, there should be a positive relationship between the high SF peak response rate and the average rate of change of $f_{\mathrm{fit}}$. Furthermore, if high SF response rate is also a function of the stability of $\sigma_{fit}$, there should be a negative relationship between the high SF peak response rate and the average rate of change of $\sigma_{fit}$. To test if this was the case, I calculated the average rates of change for $f_{\mathrm{fit}}$ and $\sigma_{fit}$ in all units ($\frac{1}{T}\Delta f$ and $\frac{1}{T}\Delta\sigma$ respectively, where T is the total number of time steps in each respective function; Δf and Δσ is the difference between initial and final values). I then calculated the high SF peak response rate in each unit as the peak response to SFs ranging from 0.18 – 0.32 cpd. However, I found no significant correlation between the high SF peak response rate and average rates of change for $f$ or σ.

In summary, the LNP model was able to generate responses resembling a large repertoire of responses observed from recorded units. Furthermore, STRFs from recorded units showed changes over short timescales which reflect coarse-to-fine dynamics. However, STRFs from recorded units did not resemble the simulated STRFs which produced coarse-to-fine dynamics in the LNP model. Therefore, the simulated STRFs which were able to generate coarse-to-fine-like responses are likely not biologically plausible. More modeling paradigms should be explored to determine whether different simulated STRFs can reproduce the responses observed from recording data.

# 5. Discussion

In this study, I set out to determine whether a purely feedforward model could explain the diversity of coarse-to-fine responses observed from recording data. Interpretations of the results are made given the following assumptions: 1) There must exist some STRF dynamics which can produce coarse-to-fine-like responses; 2) Any STRF dynamics which can produce coarse-to-fine-like responses must be biologically plausible; 3) Recorded units with coarse-to-fine responses must also have coarse-to-fine STRF dynamics. Results from the model demonstrate that coarse-to-fine responses can indeed be produced with a feedforward model. Furthermore, STRFs mapped from recorded units with coarse-to-fine responses had a SF component that increased over time, demonstrating that coarse-to-fine dynamics are indeed observable in the STRF. However, STRFs mapped from recorded units did not resemble simulated STRFs which were able to generate coarse-to-fine-like responses. Given the starting assumptions, it is therefore unlikely that a purely feedforward mechanism can explain responses observed in the recording data because the simulated STRFs that produced coarse-to-fine-like responses are not biologically plausible. However, responses modeled with STRFs resembling those observed in recorded units produced simulated response profiles closely resembling the early portion of the recorded response. This finding suggests that coarse-to-fine responses occur because of a mechanism with both a feedforward and feedback component.

## 5.1) Coarse-to-fine response profile diversity

The results show that V1 generally processes SF information in a temporally dynamic fashion in which SF preference evolves from low SF to high SF over short times scales. Furthermore, a wide variety of response profiles was observed with respect to the firing rates of low SF vs. high

SF responses. The full repertoire of response profiles was typically dominated by units that had an enhanced low SF response rate. However, many units fell outside of this typical presentation, with units that had nearly equal low SF and high SF response rates and units that had enhanced high SF response rates. Additionally, the results show that the temporal progression of coarse-to-fine responses occurs in a smooth, continuous fashion rather than an abrupt discontinuous fashion. Because of this, many of the units recorded showed responses which had a coarse-to-fine temporal progression that was limited to the lower band of SFs. In essence, these “low bandwidth” coarse-to-fine cells responded to each progressively higher SF with a progressively higher time lag but did not respond to SFs above 0.18cpd. Furthermore, a small subset of the units recorded showed responses that were either excitatory or suppressive but had no temporal dynamics with respect to SF.

A recent study exploring coarse-to-fine processing (Skyberg, R., Tanabe, S., Chen, H. & Cang, J., 2022) demonstrates both similarities and differences to the results obtained from this study. The main similarities between both studies include the observation that SF preference in V1 evolves over short timescales from low SF to high SF. Furthermore, a wide variety of response profiles with respect low SF vs. high SF firing rate was observed. However, the main difference is that coarse-to-fine responses in this study occurred in a smooth continuous fashion rather than a discrete and abrupt fashion. This difference in results may have occurred due to the difference in stimulus sets used to trigger the responses. In this study, the SFs of the gratings were defined as a harmonic series of the fundamental frequency (0.02cpd) which ranged from 0.02cpd to 0.32cpd. The decision to use this set of SFs was made to enable mapping of the STRF from the neuronal responses, however the regular step size between SFs may explain why smooth responses were observed. Initially, a stimulus set identical to the one used in Skyberg, R., Tanabe, S., Chen, H. & Cang, J., 2022 was used in the current study (results not shown). Stimulus triggered responses acquired with this stimulus set also had abrupt and discrete responses, lending support to the idea that the difference in responses between both studies is due to the difference in stimulus sets used.

### 5.2) Interpretations of the LNP modeling results

The results from the LNP model show that the variation of two parameters of the simulated STRFs can produce results that resemble a large portion of the responses observed from recorded units. These parameters are the frequency ($f$) and diameter ($\sigma$) of the Gabor filters used to simulate the STRFs in the model. The model is set up such that $f$ and $\sigma$ evolve in an opposite manner over time where $f$ increases as $\sigma$ decreases. The relative rates of the low SF and high SF responses from the model depend mainly on the rates of change of $f$ and $\sigma$. Furthermore, the way these parameters co-evolve affects the band of SFs to which the model is responsive. As the initial rate of change of $f$ increases, the response to high SFs increases and a response to a wider band of high SFs is observed. However, as the rate of change of $\sigma$ decreases in concert with $f$, the response to high SFs becomes suppressed. In extreme cases where the initial rate of change of $\sigma$ is very high, responses to high SFs barely rise above baseline and the overall response resembles a low bandwidth coarse-to-fine unit response. Although the model was able to capture a wide range of responses observed in recorded data, it was not able produce a fully suppressive response and it was not able to produce suppression of low SFs during the high SF response. Further experiments will explore whether varying different STRF parameters or using complex STRFs can produce the response properties that were not observed in this study.

Although the $f$ parameter influenced both the response rates and the band of SFs to which responses were observed, it had a larger contribution to the bandwidth of the response. Conversely, the main influence of $\sigma$ appeared to be the relative rates of the low SF vs high SF responses. Indeed, in condition 2, where $\sigma$ remained constant, the rates of the low SF vs high SF responses were nearly identical. Furthermore, in condition 3, where $f$ remained constant, responses were only observed to low SFs. These results demonstrate that coarse-to-fine responses may depend on two distinct aspects of the STRF. The first is the overall diameter of the STRF which mainly determines how the response rate changes over time. The second is the width of each excitatory and suppressive STRF subfield in the direction orthogonal to its orientation (parameterized as frequency), which mainly determines how the SF preference of the unit evolves over time.

### 5.3) Comparison of simulated STRFs and recorded unit STRFs

STRFs mapped from recorded units had dynamics which mostly differed from the simulated STRFs that were able to produce coarse-to-fine-like responses. Results from the model suggest that the $f$ parameter must increase over a wide range of SFs, at least 0.04-0.16cpd, for a response to be observed at a wide enough band of spatial frequencies such that there is a low SF and high SF peak. Furthermore, the model also suggests that $f$ must have a higher rate of change than $\sigma$ to overcome suppression from the decreasing diameter of the STRF. In most cases, mapped STRFs had $f_{fit}$ values that only increased from 0.03-0.06cpd. Furthermore, $f_{fit}$ typically increased at an equal rate to which $\sigma_{fit}$ decreased; therefore, the rate of change of $f_{fit}$ was not fast enough to overcome the suppressive contribution of decreasing $\sigma_{fit}$ values. These results suggest that the feedforward pathway contains high SF information, but that information is progressively suppressed along the pathway. Indeed, attempts to model responses with parameters retrieved from the fits to the mapped STRFs produced coarse-to-fine-like responses which were limited to the lower bound of SFs. Therefore, a potential explanation behind coarse-to-fine processing is that it begins as a feedforward process where high SF information is progressively suppressed and then subsequently enhanced/recovered through a different mechanism.

The question of what mechanisms could enhance the neuronal response to high SFs remains to be answered. A potential explanation for this could be feedback input from higher visual areas. This feedback input, specifically stemming from area LM, has been shown to be a driving factor in creating a second receptive field in excitatory V1 neurons (Keller, A. J., Roth, M. M. & Scanziani, M., 2020). This so-called feedback receptive field (fbRF) is primarily driven by stimuli outside of the classical feedforward receptive field (ffRF). Stimulation of fbRFs generates response curves that highly resemble the high SF portion of a coarse-to-fine response. That is, responses elicited via fbRF stimulation are temporally delayed and tend to have a lower amplitude than responses elicited via ffRF stimulation. Importantly, responses elicited via co-activation of the fbRF and ffRF are antagonistic when stimuli are uniform in space and synergistic when stimuli are nonuniform in space. Coarse-to-fine responses have been observed to naturalistic images (Skyberg, R., Tanabe, S., Chen, H. & Cang, J., 2022) which are nonuniform in space; therefore, responses to these stimuli may be driven by co-activation of the fbRF and ffRF. It remains to be seen whether stimuli that are nonuniform in time lead to synergistic fbRF and ffRF driven responses; if this is the case, responses triggered by 60hz random gratings may also be driven by fbRF and ffRF co-activation.

If coarse-to-fine processing does indeed elicit co-activation of the fbRF and ffRF, then the enhancement of the high SF response in V1 cells could be due to a convergence feedback mechanism. In essence, several V1 cells with low rate high SF responses may converge onto a single neuron in area LM. The convergence of activity from many V1 cells would result in an area LM response with higher SF tuning across the board. Feedback projections to V1 would therefore increase the firing rate of the entire V1 response, thus raising the rate of high SF responses significantly above the threshold firing rate. A quick way to provide support for this idea would be to record V1 responses to 60hz random gratings while an animal is under anesthesia. Because anesthesia affects HVAs to a higher degree than V1 (Lamme, V. A. F., Zipser, K. & Spekreijse, H., 1998), one would expect anesthesia to suppress the high SF portion of V1 responses to a such a degree that they are indistinguishable from baseline firing. This has indeed been shown to be the case (Skyberg, R., Tanabe, S., Chen, H. & Cang, J., 2022), further supporting the hypothesis that feedback from HVA may contribute to coarse-to-fine processing. However, further experiments must be done to test how and to what degree HVAs contribute to coarse-to-fine processing.

## 5.4) Future Directions

The main question raised by the results of this study is whether the high SF portion of the V1 response is enhanced via feedback from HVAs. Therefore, a preliminary analysis to provide support for this hypothesis would be to determine whether high SF response rates change as a function of V1 layers. Because all V1 layers except layer 4 (the main dLGN input layer) receive input strong input from HVAs (Keller, A. J., Roth, M. M. & Scanziani, M., 2020), the expectation is that units originating from these layers will have enhanced high SF response rates. Units recorded from V1 can be sorted by layer via current sink density (CSD) analysis. I will therefore use CSD analysis to separate my data into two populations: input layer units (ILUs) and non-input layer units (NILUs). I will subsequently conduct a statistical test to determine whether the mean high SF response amplitudes between the populations significantly differ. If the mean high SF response rate of the NILU group is significantly higher, this would provide good preliminary evidence that HVAs are involved in enhancing the high SF portion of the response.

Beyond this preliminary analysis, further experiments should be done to link HVA feedback to coarse-to-fine processing. An informative experiment would be to silence individual HVAs while simultaneously recording from V1. This can be done by expressing channelrhodopsin (ChR2) to HVAs via viral injection and targeting individual HVAs with a laser to optogenetically activate inhibitory neurons. This protocol has been successfully implemented previously with minimal direct impact of scattered laser light on V1 (Keller, A. J., Roth, M. M. & Scanziani, M., 2020). After successfully expressing ChR2, I would record responses of V1 neurons to several repeats of each stimulus condition with and without optogenetic silencing of an individual HVA. From this experiment, I expect the same single units within the NILU group to have suppressed high SF response rates during silencing. Furthermore, I would expect no change in units within the ILU group between silenced and unsilenced conditions. Finally, I would repeat this experiment for each individual HVA to determine which HVA has the highest impact on high SF response rates in V1.

Upon identifying the HVA with the highest impact on high SF response rates in V1, I would conduct simultaneous extracellular recordings in V1 and the HVA during visual stimulus presentation. By conducting a simultaneous recording, I hope to establish a final link between HVA feedback and high SF enhancement in coarse-to-fine processing. This link would be established by conducting an analysis of the time course of the response in V1 and the HVA. Essentially, if HVA feedback is truly enhancing the high SF portion of the V1 response, the peak of the response in the HVA must occur prior to the peak of the high SF response and follow the peak of the low SF response in V1.

Beyond investigating potential HVA mechanisms contributing to coarse-to-fine processing, future experiments must also include analysis of complex STRFs. This may require a more direct method of estimating STRF parameters because complex STRFs are composed of multiple simple Gabor-like STRFs. Alternatively, complex STRFs can be separated into corresponding simple components; under this protocol, the fitting procedure used in this study can be applied to each individual simple component of the STRF. In either case, a full picture of the mechanisms facilitating coarse-to-fine processing will require analysis of both simple and complex cells. Furthermore, future experiments should also explore potential contributions of inhibitory cell types in V1, however it is not currently clear how this can be achieved with resources available. One could potentially mark and record inhibitory cell types in V1 via optotagging, however previous attempts of cortical optotagging have typically netted a low yield. Nonetheless, a feasible analysis would be to look at the temporal SF dynamics of broad vs.

narrow spiking cells. Although narrow spiking is a nonspecific trait of inhibitory neurons in V1, characterizing their dynamics could still provide an interesting preliminary picture of how inhibitory neurons process SF.

## 5.5) Limitations

A major limitation to this study is the size of the dataset used for analysis. The full dataset includes a total of 72 single units recorded from 11 mice. The low yield of single units is a consequence of ongoing troubleshooting of the electrophysiological recording procedure. Results from section 4.3 are made with a dataset of only 22 units because analysis of complex cells is not included. Furthermore, the cells which exhibited a coarse-to-fine response tended to have very low high SF response rates. Because of this, analysis of the correlation between $f_{fit}$ and $\sigma_{fit}$ average rates of change and peak high SF response may be unreliable.

Another limitation is that the modeling results are still preliminary. Therefore, much of the interpretations of the modeling results are made via qualitative assessment rather than statistical analysis. Furthermore, the extent to which the parameter space of the model was explored is limited. Therefore, a more systematic exploration of the parameter space of the model is needed. Additionally, the response of the model is generated at a 60hz temporal resolution whereas recorded responses are binned at 1ms intervals. Therefore, the model may be missing dynamics only present at higher temporal resolutions.

Finally, the estimation of the parameters from STRFs mapped from recorded units was achieved with nonlinear curve fitting. This method may be prone to biases which arise due to the minimization algorithm falling into local minima that do not capture the true parameters of mapped STRFs. Therefore, a more direct method for estimating the STRF parameters is needed.

# 6. References


1. Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383. https://doi.org/10.1016/0010-0285(77)90012-3
2. Lamme, V. A. F., Zipser, K. & Spekreijse, H. (1998). Figure-ground activity in primary visual cortex is suppressed by anesthesia. *Proceedings of the National Academy of Sciences*, *95*(6), 3263–3268. https://doi.org/10.1073/pnas.95.6.3263
3. Sceniak, M. P., Ringach, D. L., Hawken, M. J. & Shapley, R. (1999). Contrast's effect on spatial summation by macaque V1 neurons. *Nature Neuroscience*, *2*(8), 733–739. https://doi.org/10.1038/11197
4. Bredfeldt, C. E. & Ringach, D. L. (2002). Dynamics of Spatial Frequency Tuning in Macaque V1. *The Journal of Neuroscience*, *22*(5), 1976–1984. https://doi.org/10.1523/jneurosci.22-05-01976.2002
5. Suder, K., Funke, K., Zhao, Y., Kerscher, N., Wennekers, T. & Wörgötter, F. (2002). Spatial dynamics of receptive fields in cat primary visual cortex related to the temporal structure of thalamocortical feedforward activity. *Experimental Brain Research*, *144*(4), 430–444. https://doi.org/10.1007/s00221-002-1061-5
6. Singer, S., & Nelder, J. (2009). Nelder-mead algorithm. *Scholarpedia*, *4*(7), 2928.
7. Einevoll, G. T., Jurkus, P. & Heggelund, P. (2011). Coarse-to-Fine Changes of Receptive Fields in Lateral Geniculate Nucleus Have a Transient and a Sustained Component That Depend on Distinct Mechanisms. *PLoS ONE*, *6*(9), e24523. https://doi.org/10.1371/journal.pone.0024523
8. Purushothaman, G., Chen, X., Yampolsky, D. & Casagrande, V. A. (2014). Neural mechanisms of coarse-to-fine discrimination in the visual cortex. *Journal of Neurophysiology*, *112*(11), 2822–2833. https://doi.org/10.1152/jn.00612.2013

9. Chung, J. E., Magland, J. F., Barnett, A. H., Tolosa, V. M., Tooker, A. C., Lee, K. Y., Shah, K. G., Felix, S. H., Frank, L. M. & Greengard, L. F. (2017). A Fully Automated Approach to Spike Sorting. *Neuron*, *95*(6), 1381-1394.e6. https://doi.org/10.1016/j.neuron.2017.08.030
10. Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M. & Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. *Nature*, *571*(7765), 361–365. https://doi.org/10.1038/s41586-019-1346-5
11. Yang, L., Lee, K., Villagracia, J., and Masmanidis, S.C. (2020). Open source sil-icon microprobes for high throughput neural recording. J. Neural Eng. 17,016036.
12. Keller, A. J., Roth, M. M. & Scanziani, M. (2020). Feedback generates a second receptive field in neurons of the visual cortex. *Nature*, *582*(7813), 545–549. https://doi.org/10.1038/s41586-020-2319-4
13. Tanaka, H. & Sawada, R. (2022). Dynamics and Mechanisms of Contrast-Dependent Modulation of Spatial-Frequency Tuning in the Early Visual Cortex. *The Journal of Neuroscience*, *42*(37), 7047–7059. https://doi.org/10.1523/jneurosci.2086-21.2022
14. Skyberg, R., Tanabe, S., Chen, H. & Cang, J. (2022). Coarse-to-fine processing drives the efficient coding of natural scenes in mouse visual cortex. *Cell Reports*, *38*(13), 110606. https://doi.org/10.1016/j.celrep.2022.110606