1 TO DO

1. *Include Mikes kitchen model. Stored on Github at https://github.com/michaelmike123/V-rep.git. File format not regular VREP scene. Open on a windows machine and export to a regular VREP scene.*
2. *Points in italic need review from Bryan.*
3. *Some of the images are low quality - from old presentations. Original images for these need to be recaptured.*
4. *Include reference to code: https://github.com/salkhan23/ITCortex.*

2 Introduction

- IT Cortex background
  - Objects in our natural visual experiences exhibit a variety of transformations - they can fall anywhere on the retina, appear at different distances, orientations, illuminations, occlusions, deformations, colors, outlines and often coexist with multiple objects. Despite these challenging viewing conditions, the brain can rapidly and effortlessly decode object identity. In the primate brain, visual information processing ~~is known to~~ progresses along two pathways: the ventral 'what' stream and the dorsal 'where' stream. The Inferior Temporal (IT) cortex, located at the top end of the ventral stream, is thought to be responsible for object detection and recognition. Feed-forward input to the IT cortex consists of spikes from neurons in lower ventral stream layers. Given the variety of transformations objects display in complex environments and the variability in spike patterns of lower layer neurons, each IT object encounter can be considered unique, [2]. Yet, object identity has been shown to be decodable from population responses of IT neurons using simple biologically plausible classifiers, [5].

    Add references here and be careful with wording e.g. "contribute to" would be better than "responsible for"

  - This remarkable capability of the IT cortex was thought to extend to individual IT neurons as well. However, responses of individual IT neurons were found to be strongly impacted by these transformations. This lead to the contemporary view that the IT cortex, as a whole, encodes not only object identity but their attributes as well. How the IT cortex mechanistically achieves this is still not understood. Clearly there is a need to develop a strong understanding of how these transformations influence IT neuronal responses at both the individual and population level.

    Ref

This stuff mostly belongs in methods

- Background on specific problem
  - *Tuning curves in the literature*
    - Often papers isolate one or two stimulus dimensions and present tuning profiles as stimuli change along the chosen dimension(s).
    - *Why collect data from multiple tuning curves and combine into a single model.*
    - *Similar work: find articles that attempt to do this. If cant find any, why not. Drawbacks of approach. Why haven't people attempted this before?*
  - *Why we can consider independent tuning curves and use the product to get net firing rate*

    This is an assumption we use when there is no data to the contrary. Another way to put this is we assume multidimensional tuning curves are separable.

  - *Consider only inanimate objects*
    - We consider only inanimate objects. Localized clusters for animate objects have been identified in the IT cortex, while processing of inanimate objects appears to be more widely distributed. - reference
    - ~~It has also been shown that~~ IT neuronal response properties differ for these two categories. - in what way? is this relevant?
  - *Single tuning profile for all objects*
    - All objects share a common tuning profile for each considered transformation.

      I don't know what this means

        * [4] found that IT neurons were arranged in columnar structures that shared similarities in their preferences. We uses this as justification for a single tuning profile for all objects.
        * In [18, 6] it was found that object preference at least for the neurons most preferred object was mostly preserved under various transformations. Although they did notice some deviations from this general rule.
    - *What about noise correlations between nearby neurons. Why can we ignore these?* yes - if you have a reference you can briefly note a limitation in the discuss
    - *What about spatial arrangement of IT neurons. Our approach is to use statistical population properties rather than detailed spatial maps. Why not use spatial maps? drawbacks? problems? Why is this a good compromise?* this is just a thing you haven't done (one of many)
    - *We do not consider inhibitory and excitatory neurons separately. Why?* - are they distinguished in the paper you draw from? if not, you're stuck.
- Comaprison with other IT models:
    - Other models focus on object detection and classification capability of the IT cortex.
    - Our emphasis is on modeling tuning properties and realistic firing patterns of IT neurons.
    - *How does our model ~~help/assist~~* complement *other IT cortex models?* - as input to ~~z~~ models of other areas like parahippocampal areas, AIP
- Specific Objective
    - Develop an integrated model that can replicate observed individual and population responses of the IT cortex given object ground truth in a scene. Specifically, we target: (1) object selectivity, (2) maximum fire rates, (3) position, (4) size, (5) orientation, (6) occlusion, (7) clutter and (8) dynamic responses of IT neurons.
        * *Why have we chosen these specific transformations.* don't ask me - you chose them
        * *What about the others that have been left out. Is the plan to include those later?*
    - Motivation for doing so (overall goal) - once we build models with this IT model as input, we can test sensitivity of various funct (e.g. object recognition) to IT statistical properties, by varying this mod
- Methodology:
    - For each tolerance, find statistical fit(s) for observed tuning profiles.
    - For each statistical model develop models for distributions of their parameter(s) and reproduce realistic population level responses.
    - Use Virtual Robot Experimentation Platform (V-REP) to generate a complex visual scene and develop techniques for extracting and quantifying relevant ground truth (identity + attributes) from scenes.

3 Methods

3.1 Selection Criteria for Literature Results to model

- *Chose results that provided the highest granularity and most extensive results for the stimulus dimension of interest.*
- *Explain techniques used to align results from different papers, as some properties are quantified differently.*
- *Data from references was extracted using WebPlotDigitier [1]. Fix this reference.*

under wrong heading

3.2 Tuning Profiles

3.2.1 Stimulus Selectivity

- It is well established that IT neurons respond to stimuli more complex then those of lower layers [16].
- Neurophysiological studies typically test IT neurons with a range of different objects across a broad range of object categories.
- ~~It has prooven difficult to come up a single consistent quantification of object identity.~~
- A common characterization of object identity, that goes around having to identify its dimensionality, is the percentage of objects a neuron responds with above background activity levels among the total stimuli set, objects it is selective for.
- [9, 18] both recorded a wide range of object selectivities using different metrics.
- We quantified selectivity over objects of model neurons by excess Kurtosis - a measure of the peak sharpness and tail heaviness of a distribution. A distribution with a high Kurtosis generates large responses (peak sharpness) for a few object while most objects generate minimal respones (tail heaviness). While a distribution with a low Kurtosis responds similarly to several objects (broad peaks) while no discernible responses for the rest (small tails). – give the reference and explain you're using it because it's a large dataset

If this is common, (1) don't bother justifying it and (2) give lots of references

$$SI_K = \frac{\sum_{i=1}^{N}(r_i - \bar{r})^4}{Ns^4} - 3 \tag{1}$$

Where

r_i is the response of the neuron to the ith test stimulus,

$\bar{r}$ average response of the neuron across all stimuli,

s is the standard deviation of the responses,

N is the number stimuli.

In case of population sparseness,

r_i is the response to the ith model neuron,

$\bar{r}$ average response to the test stimulus,

s is the standard deviation of responses to the test stimulus,

N is the number of neurons.

too redundant just use this version

- Population sparseness, population responses to a specific stimulus, was also quantified by excess Kurtosis. It was found in [9] that population spareness was typically higher than neuronal selectivity. Hence a visible object typically generates large responses from a few neurons, while individually, neurons respond to several objects.
- [9] also observed that population average response to individual test stimuli were similar but different neurons responded well to different stimuli. They concluded that object identities are encoded in the activity patterns of the population and not by mean population responses. *(Can get figure for this result - plot of average fire rates for different objects and multiple plots of firing rates of individual neurons to diffrent stimuli.)*
- Similar to [9], neuronal selectivities over objects are modeled as Gamma probability distributions.
- The authors justify the selection of a gamma distribution with different scale and shape parameters to create a statistically inhomogenous population of neurons which they show is necessary to achieve the result population sparseness is greater than individual neuronal selectivity.

- Distributions for shape and scale parameters for the gamma selectivity profile were also modeled as Gamma distributions. Parameters for their distributions were based on values from [9] but were optimized to account for non-optimal position and size of their test stimuli.
    - non-optimal position: all stimuli were presented foveally. However the receptive field centers of IT neurons, their optimal positions, are distributed around the fovea. To account for this we created a large sample of position profiles and determine the amount of deviation expected if stimuli are presented at (0, 0) to all neurons.
    - non-optimal size: The largest dimension of all test stimuli extended 7° degrees. To account for this we generate a large sample of size tuning profiles of IT neurons and determine the amount of deviation expected if stimuli are presented with a size of 7° degrees.
    - Finally the mean and variance of these deviations were calculated and used to adjust the shape and scale distribution parameters. *Add more details on how the mean and variance of the deviations was used and what is adjusted because of them [Bryan].*
    - *Also what is the impact of this modified distributions? lower Kurtosis across the population compared to Lehky Kurtosis distribution?*
- A second measure of neuron selectivity, the activity fraction [18], was also calculated for each model neuron.

$$SI_{AF} = \frac{(N-1)}{N}\left(1 - \frac{\left(\sum \frac{r_i}{N}\right)^2}{\sum \frac{r_i^2}{N}}\right) \tag{2}$$

where

r_i is the response to the ith stimulus

N is the number of test stimuli.

- Activity fractions range between [0, 1]. A neuron with an activity fraction of 0 responds equally to all stimuli (less selective) while a neuron with a selectivity index of 1 shows preference to a few stimuli only (highly selective).
- Neuron selectivity assessed as the activity fraction was used to model decreases in tolerances as selectivity increases in several tuning profile as observed in [18]. Maybe better to bring this up when you get to tolerance
- At initialization, a list of all objects in the scene is passed into each model neuron. The neuron ~~randomly~~ uniform random assigns a value between (0, 1) to each object. We use the inverse of the cumulative distribution function (CDF) of the neuron's selectivity profile normalized by its maximum value to find the neurons preference for each object.

3.2.2 Maximum Firing Rate

- The selectivity profile of model neurons was also used to generate their maximum firing rates.
- The inverse CDF was used to determine the unnormalized selectivity at which the CDF is 0.99.
- This was used as the maximum firing rate of model neurons.
- *Other then the convenience of using the selectivity distribution for deriving the maximum fire rate distribution, need to come up with a model that models the literature or some property of IT neurons.*
    - *Candidate: [9] found a strong negative correlation between an IT neurons selectivity (Kurtosis) and their mean response across all stimuli. As a neuron becomes more selective its mean firing rate decreases, Figure 5Ai. Our current maximum firing rate model (selectivity where CDF=0.99) does not clearly model this observation. A highly selective neuron will have a gamma distribution with a high Kurtosis. Because high Kurtosis distributions have long tails, its selectivity at CDF=0.99*

will also be a large. Hence, the more selective a neuron, higher is maximum firing rate. Although a highly selective neuron may still have a low average firing rate, because a few neurons return responses close to maximum and the average is taken over all the objects. The average response of less selective neurons may also be low - several stimuli have similar responses, but the majority have no responses. Maybe average firing rate verses selectivity is not a good measure. Somehow we need to model this result in a clearer way.

- *Currently the overall model is showing low firing rates, possibly due to the our max fire rate model.*
- *Need to plot the average firing rate versus Kurtosis selectivity and see if our model agrees with this result.*

3.2.3 Position Tuning

- Spatial receptive fields were modeled as normalized 2D Gaussian distributions with parameters, receptive field center and position tolerance. - also give gaussian parameters in terms of these
- [14] found that receptive field centers, preferred object locations, were distributed around the fovea with a slight bias towards the contralateral visual hemisphere. Coordinates (X, Y) of the receptive field center were modeled as Gaussian distributions. Parameters of these distribution were estimated by least square error (LSE) fitting of the receptive field center of masses of [14]. x $\sim N(1.82, 2.02)$, y $\sim N(0.62, 2.12)$. Where x, y are in degrees of eccentricity from the fovea. - give figure ref.
- The position tolerance, the extent of the spatial receptive field was quantified as twice the standard deviation of the underlying Gaussian distribution. why is this from 14?
- [14, 18] both observed a broad range of position tolerances across the population.
- Spatial extent in the x and y direction of receptive fields were found to be similar in [14] and we use a single position tolerance for both.
- A Gamma distribution was used to model the position tolerance distribution. To find the parameters of this distribution first we did a linear regression fit of the position tolerance versus neuron selectivity (activity fraction) scatter plot of [18]. This was used to model how mean position tolerance decreases with activity fraction selectivity. Second we assumed the shape parameter of the distribution was fixed and the scale parameter, derived from the mean, decreased with selectivity. Finally we used maximum likelihood estimation to find the best fit shape and scale parameters subject to the above constraints. This models two effects seen in the data: average position tolerance decreased with selectivity and variance from the mean decreases with selectivity [18].
- The normalized position firing rate of is given by:

$$r_{position}(x, y) = \exp\left(\frac{-\left((x - x_c)^2 + (y - y_c)^2\right)}{2\sigma_{PT}}\right) \tag{3}$$

where

x, y are the position coordinates of the stimulus in the x and y in radians of eccentricity,

x_c, y_c is the receptive field center of postion tuning profile,

σ_{PT} is the position tolerance of the neuron in radians of eccentricity.

- For each model neuron the spatial extent of the receptive field was also quantified by the square root of its area. Here the area of the receptive field was defined by a circle with radius equal to the position tolerance. *Currently we use half the position tolerance to define the area. Change to position tolerance only. Our size tuning model uses this to determine the maximum size of stimulus that it can respond to.* did you use this for anything? what?
- *Because we cannot generate activity fraction > 0.6, in our model by using the activity fraction to determine the receptive field size, the model may not be able to generate some of the larger receptive fields seen in the IT cortex by [18]. Need to investigate further.*

In each section, first say succinctly what you did, and then fill in the more minor details. These sections should not start like mini-introductions. Say what you did, then briefly refer to literature to justify

3.2.4 Size

- Size tolerance of model neurons was modeled as lognormal distributions with parameters preferred stimulus size and size bandwidth. *Why is this profile log normally distributed but other tuning profiles are symmetric on the linear scale.* Maybe because retinal size varies inversely with distance to object, which is maybe ~ uniform in nature – but you shouldn't theorize in the paper. – just fit the data.
- Stimulus size was quantified as the distance between the outer edges along the longest axis of the stimulus.
- [6] found a wide range of preferred stimulus sizes and found little correlation with the neurons spatial receptive field size. Although preferred stimulus sizes were generally smaller.
- A lognormal distribution was used to model preferred stimulus sizes. Across their population of IT neurons [6] found two peaks, at preferred stimulus size 3.4 and 27 degrees. However, we ignored preferred sizes that were greater than or equal to 27°. These sizes were close to the largest stimulus size and several neurons had receptive field sizes that could accommodate much larger stimuli. It is likely that the preferred stimulus size of these neurons would be different if larger stimuli were used. This is not to say that preferred sizes cannot be greater than 27° are not generated, but that the frequency of their occurrence follows the distribution and trends seen at lower sizes. Parameters for the preferred stimulus sizes distribution were then determined by maximum likelihood fitting of the remaining data.

Separate section earlier on maximum likelihood fitting details

- Size bandwidth was quantified as the distance between the upper and lower half magnitude responses of a neurons size tuning profile in octaves.
- Little correlation was found between a neurons size bandwidth and its receptive field size.
- A lognormal distribution was also used to model the distribution of size bandwidths. Parameters for this distribution were determined by maximum likelihood fitting of size bandwidths measured in [6].
- Parameters of the lognormal size tuning profile are derived from a neuron's preferred stimulus size and size bandwidths as

$$\mu_s = log_2(preferred\,stimulus\,size) \tag{4}$$

$$\sigma_{ST} = \frac{size\,bandwidth}{\sqrt{2\ln(2)}} \tag{5}$$

need symbols for these

why?

- The normalized size firing rate is given by:

$$r_{size}(s) = \begin{cases} \exp\left(\frac{-(\log_2(s)-\mu_s)^2}{2\sigma_{ST}^2}\right) & \text{if } s \leq s_{max} \\ 0 & otherwise \end{cases} \tag{6}$$

where

s is the size of the stimulus defined as the maximum distance between the outer edges of the stimulus along the x, y, or z dimension in radians of eccentricity,

μ_s is the preferred stimulus size in radians of eccentricity,

σ_{ST} is the size tolerance of the neuron in radians of eccentricity.

$s_{max} = 2\sigma_{PT}$ is the maximum stimulus size that can fit within the receptive field of the neuron.

~~Note that we model the distribution as a normal distribution by taking the logarithm of the input stimulus size.~~

- The spatial receptive field defined in terms of its areal extent was used to determine a maximum stimulus size, its diameter, for each neuron. Responses to all stimuli of size above this maximum were masked. *Should this be taken out/modified? (1) There should be a gradual decrease in firing rate as the stimulus increased to a size greater than the maximum, rather then a sharp cutoff. Diffrent mask? (2) can we say that recorded tuning curves of this type were not completely measured.*

what do you mean by masked?

There should be a gradual decrease, just like in the neurons for which larger-than-optimal stimuli were presented.

- *A negative correlation between activity fraction selectivity and size tolerance was observed by [18]. Size tolerance was quantified as the drop in firing rate as the size changes from the optimal size of a neuron. Currently this is not included in the model, it would be useful to include as done for position tolerance. Although the range of sizes test (1, 2, 4, and 6°) is much smaller than the range of sizes tested in [14].*

3.2.5 Orientation

This is brief and interesting enough that you can leave it at the beginning of 3.2.5. It's clear that you are about to explain how you model this.

- [12] found that IT neurons typically had a preferred view of an object - most familiar view - and their responses gradually declined as views rotated away.
- This preference persisted even as the monkey gained familiarity with the object and developed view invariant recognition at the behavioral level.
- Individual IT neurons developed preferences for different views.
- We model the ~~base~~ rotation tuning profile of IT neurons as a single Gaussian distribution with parameters preferred view and rotational tolerance.
- [12] also found that the tuning curves of some neurons were centered around two views. In all such cases, peak views displayed some level of symmetry, including mirror symmetry. Test objects were irregular wire or amoeboid objects. Because the objects were never completely symmetrical, their bi-modal tuning curves had amplitude differences between peaks.
- We do not model level of symmetry between views and use a binary metric to judge similarity. All symmetric views share the same peak amplitude and we do not model amplitude differences between peaks.
- A few neurons also displayed view invariant responses to all object views. A sample view invariant tuning curve is shown in [12], however the object is not shown and it cannot be determined whether the symmetry of the object contributed to this view invariance.
- We assume multiple peaks in tuning curves are the result of symmetries in views of the object. From a neurons perspective, there is still a single preferred view and responses to views is based on how similar the view is to its most preferred view rather than absolute orientation.
- To model symmetric objects, we assume each object has ~~defines~~ a period of symmetry around the x, y, and z axis. Where the symmetry period is defined as the number of views within a 360° rotation about the target axis that are identical to the preferred view. ← that would be frequency - period is in ° or rad.
- Within the symmetry period each object view is distinguishable as a deviation from a model neurons preferred orientation. While an orientation outside maps to a view within the symmetry period. – Pair this with equation
- Mirror symmetric objects are modeled as the sum of two Gaussian distributions with the second Gaussian centered around the mirror symmetric preferred view and with equal rotational tolerance.
- Each neuron then has a single rotation tuning curve that ranges from $(-\pi, \pi)$ for all objects it is selective for. However if an individual object defines a period of symmetry it will only see a partial view of the complete tuning profile within the defined period of symmetry about its preferred view. We assume symmetric tuning of orientations away from the preferred view in both directions.
- The normalized rotation firing rate is given by:

$$r_{rotation}(\theta, p_{sym}, m_{sym}) = \exp\left(\frac{-(\theta_{adj} - \mu_r)^2}{2\sigma_{RT}^2}\right) + m_{sym}\exp\left(\frac{-(\theta_{adj} - \mu_{r'})^2}{2\sigma_{RT}^2}\right) \tag{7}$$

where:

θ is the orientation of the object about the target ? axis in radians, A figure might be in order to explain all this. Not a high priority

θ_{adj} is θ is adjusted to lie within the valid range of the symmetry period of the object. Valid range of the symmetry period is defined as $\frac{2\pi}{p_{sym}}$,
μ_{r} is neurons preferred view of the object adjusted to lie within the symmetry period,
σ_{RT} is the rotational tolerance of neuron in radians,
$m_{sym} \in \{0, 1\}$ whether the object is mirror symmetric or not,
$\mu_{r'}$ mirror reflection of the neurons preferred view adjusted to lie within the symmetric period.

- The distribution of preferred orientation was modeled as uniform random variable within $(-\pi, \pi)$.
- In [12], it was found that rotational tolerance, defined as the standard deviation of the fitted Gaussian distribution, around the x-axis and y-axis were similar and averaged 30° for both classes of tested objects. Absent any further details on distributions of rotational tolerances we model their distribution as a normal distribution with mean 30° and standard deviation of 50°. *Why was this value for standard deviation chosen.* — why indeed — seems high and will give many -ve samples
- We also only model rotation tuning around y-axis and assume rotation tolerances about the x and z axis follower similar trends with their own preferred view, tolerances and symmetries.

3.2.6 Occlusion

- Responses of IT neurons increase monotonically with object visibility [8].
- IT neurons ~~have~~ also ~~been shown to~~ respond to object parts rather than the whole and prefer certain parts over others.
- These diagnostic parts are task dependant and behaviorally relevant to the task the monkey was actively attending.
- The level of preference for diagnostic parts varied across the population [13].
- However diagnostic parts were always preferred to nondiagnostic parts.
- To quantify an individual neurons preference for diagnostic parts, [13] use the ratio of diagnostic group variance to total variance. Depending on which parts were visible, trials were grouped as diagnostic or nondiagnostic. The amount of the total variance that was explained by the between group variance quantified a neurons preference for diagnostic parts.

$$R = \frac{Var_{diagnostic\,grouping}}{Var_{total}} \tag{8}$$

$$Var_{diagnostic\,group} = \frac{1}{2}\left((\bar{f}_{diagnostic} - \bar{f})^2 + (\bar{f}_{nondiagnostic} - \bar{f})^2\right) \tag{9}$$

where
R is the preference of the neuron for diagnostic object parts. Range (0, 1),
σ_b^2 ~~$Var_{diagnostic\,group}$~~ is the variance $\boxed{\text{between}}$ diagnostic and nondiagnostic groups,
Var_{total} is the variance across all trials,
$\bar{f}$ is the mean firing rate across all trials,
$\bar{f}_{diagnostic}$ is the mean firing rate across trials where only diagnostic parts were visible.
$\bar{f}_{nondiagnostic}$ is the firing rate across trials where only nondiagnostic parts were visible.

- A ratio of 1 means that the variance between groups is high, group mean firing rates are significantly difference from the overall mean, and the neuron shows a strong preference for diagnostic parts. While a ratio of 0 means there is no difference between group variances, group mean firing rates are the same as the overall mean, and the neuron responds equally to diagnostic and nondiagnostic parts.

- We model the occlusion tuning profile of model neurons as a two-input Sigmoid with parameters w_d, w_{nd} and b. Where w_d is the weight for the percentage of diagnostic parts that are visible, w_{nd} is the weight for nondiagnostic visibility and b is a bias term. – give the equation including % visible
- w_d and w_{nd} are chosen such that they can generate the desired preference for diagnostic parts.
- We used least square errors fitting on the scatter plot of R in [13] to fit an exponential distribution.
- We then used the inverse CDF of this distribution to generate a diagnostic parts preference ratio for each neuron. *Occasionally this inverse CDF generates a value greater then one and the code aborts. Not sure why. Implementation appears to be correct.*
- This is still underconstraint to generate w_d and w_{nd} as there is no bounds on the ranges and neither is there any relation between w_d and w_{nd}. *Can we use scipy.optimize.curve_fit to directly find a set of w_d and w_{nd} without having to rely on w_c?*
- To overcome this, we looked at occlusion tuning curves of IT neurons from other references.
- In [8] tuning curves for multiple neurons to several occluded objects were recorded. However, diagnostic and non-diagnostic visibilities were not considered separately and a single visibility measure was used.
- We assume equal % visibilities for diagnostic and nondiagnostic parts and use least square error fitting to fit all tuning curves to a single input-Sigmoid with parameters w_c and b.

$$r_{occlusion}(v_c) = \sigma(w_c v_c + b) \tag{10}$$

where

$v_c = \sqrt{v_d^2 + v_{nd}^2}$ is the combined visibility of diagnostic and nondiagnostic parts of the object. If we assume equal visibilities for diagnostic and nondiagnostic parts, $v_c = \sqrt{2}v_d = \sqrt{2}v_{nd}$. Note that v_c ranges between $(0, \sqrt{2})$.

w_c is the weight along the combined visibility axis where $v_{nd} = v_d$.

$\sigma(x) = \frac{1}{1+\exp(x)}$ is the Sigmoid function. this is one kind of sigmoid (logistic). Sigmoid just means S-shaped.

- We found the mean and variance for w_c and b across all tuning curves in [8]. We then assumed separate normal distributions for the combined weight and the bias terms.
- ~~We assume~~ The three weights are related though $\sqrt{2}w_c = w_n + w_d$. *How did we derive this relation. Need to properly explain. [Bryan].*
- Given w_c, b and R for each neuron we use nonlinear optimization to determine a set of w_d and w_{nd} that can generate a between group variance of R. The nonlinear optimization technique used is the root finder function fsolve of Scipy's Optimize library where we find the roots of the function that calculates the diagnostic preference ratio given a set of w_d and $w_n$and subtracts it from the target ratio. Initial estimate of the search for parameters is half the combined weight.

mention the algorithm fsolve uses.

- Occasionally the optimization technique is unable to determine a set of w_d and w_{nd} that can generate the required preference ratio. In such case we incease w_c until a value is found that can reproduced the target ratio. *In the simulation we increase w_c in steps of 2 until for up to 100 iterations. Unfortunately if a very high ratio is required this method still fails and the program aborts. Need a better way to search for w_c so the program doesn't crash. We do not worry to much about following the distribution of w_ c or b.* – better figure and why this is – evaluate at $w_d = \sqrt{2}w_c$ & $w_{nd} = \sqrt{2}w_c$ and check that these results surround required preference ratio (must be a bug if not)
- The normalized occlusion firing rate of a neuron is then given by

$$r_{occlusion}(\mathbf{v}) = \sigma(\mathbf{v}.\mathbf{w} + b) \tag{11}$$

where

$\mathbf{v} = \begin{bmatrix} v_{nd} \\ v_d \end{bmatrix}$ is a vector of nondiagnostic and diagnostic visibility,
$w = \begin{bmatrix} w_{nd} \\ w_d \end{bmatrix}$ is a vector of diagnostic and nondiagnostic weights,
. is the dot product,
b is the bias term.

- *This method of generating occlusion tuning curves is too convoluted. It may be hard to convince why this is a good model. Need to come up with a simpler way or justify why this is a good model.*
- Responses of IT neurons are also known to take longer times to develop when objects are occluded [8, 13]. We do not model the additional latencies seen in responses of IT neurons. *Why?*

I think you can just write it more succinct with more focus on key points

3.2.7 Clutter

This much specific introduction in a Methods subsection is certainly fine

- Responses of IT neurons are know to be influenced by nearby objects.
- Response attenuation is also seen from objects which individually do not elicit responses.
- IT neuronal responses ~~have~~ are also ~~been shown to be~~ impacted by background complexity [15].
- Under limited clutter conditions, two to three objects over a plain background, [17] showed that responses of individual and populations of IT neurons to multiple objects were smaller than the linear sum of isolated responses and were much closer to the average of responses to constituent objects in isolation. Multiple averaging rules were tried in [11] and it was consistently found that all averaging models better fit recorded responses compared with models base on the linear sum of isolated respones.
- To model the clutter response of a model neuron, first its isolated response is calculated for all objects,

$$r_{isolated}(obj_m) = R_{max} * pref_{obj_m} * r_{position}(x,y) * r_{size}(s) * r_{rotation}(\theta, p_{sym}, m_{sym}) * r_{occlusion}(\mathbf{v}) \quad (12)$$

where
$ground\,truth = x, y, s, \theta, p_{sym}, m_{sym}, v_{nd}, v_d$, is the ground truth for object m,
$pref_{obj_m}$ is the neurons normalized selectivity for object m,
R_{max} is the maximum firing rate of the neuron.

- Next isolated responses are weighted by normalized position responses and averaged to get the static ? response of neuron n

$$r_{static} = \frac{\sum_m^M w_{position\,obj_m}(x,y) * r_{isolated}(obj_m)}{\sum_m^M w_{position\,obj_m}(x,y)} + d \quad (13)$$

where
$w_{position\,obj_m(x,y)} = r_{position}(x,y)$ for object m,
M is the number of objects in the scene.
d is a random variable that models the deviation from the averaging rule.

from what distribution? / does it scale with static?

- Scatter plots of joint responses to two and three objects in [17] plotted against the isolated responses were used to find d. First the deviation of joint responses from the averaged isolated responses was found. A normal distribution with mean 0 and the standard deviation measured from the deviations was used to model d.
- By weighting isolated responses with their position weights, clutter responses are impacted by any object within their receptive field regardless of whether the neuron is selective for that object in isolation or not, consistent with [ref]
- We do not model the impact of background on clutter responses. - define "background" a bit more
- *In [18], it was found that the clutter responses of IT neurons also decrease with activity fraction selectivity. Should the impact of selectivity be included in the clutter response?*

-No, just mention it in the discussion under Limitations. It won't be possible to model everything. You just have to pick a reasonable # of important things.

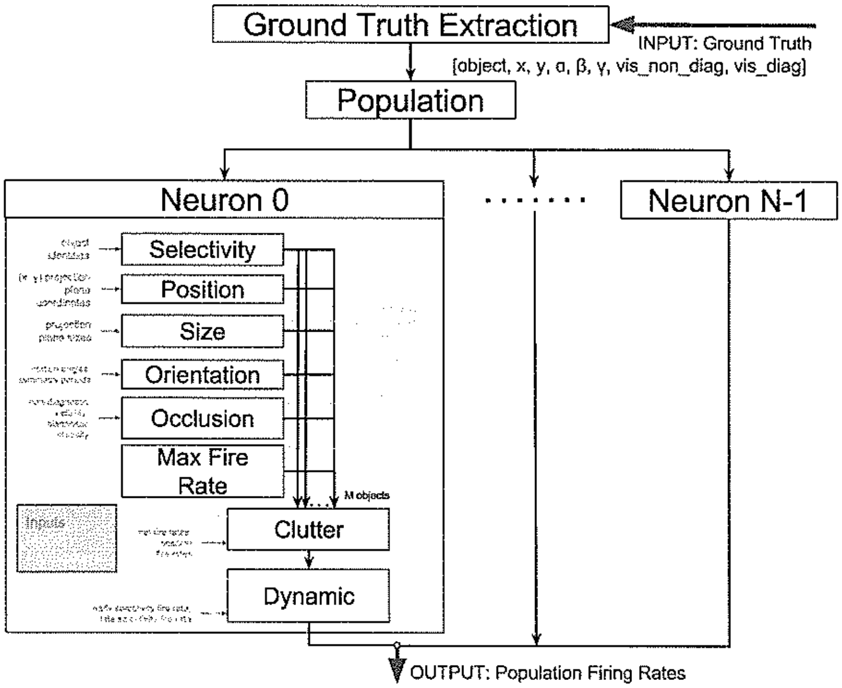


Figure 1: *Overall Model. Figure needs some updating to use symbol notations in this paper. Also we need the rotation symmetries of the objects which are not included in the ground truth above. Red and green fount items are not readable in printed document- change font & color.*

3.2.8 Dynamics

- *[Bryan]*
- $r_{net}(t) = r_{dynamic}(t) = f(r_{static}) = ?$
- *Timestep used was 5ms. But is changeable.*

3.3 Overall Model

- *Complete me*

3.4 Ground Truth Generation and Extraction

I would de-emphasize the robot. Just say,

- A test scene in Virtual Robot Experimentation Platform (VREP) [3] is used to generate ground truth for the IT cortex model.
- Several objects were placed ~~randomly~~ in the scene ~~along with the IT cortex robot~~.
- The IT cortex robot consisted of a vision sensor attached to an existing mobile robot (pioneer p3dx).
- The main job of the robot was to give mobility to the vision sensor.
- The vision sensor has an adjustable projection plane onto which it does a 2D rendering of its perspective of the 3D test scene. - do you have to say this? isn't this true of all such sensors?

The IT model was attached to a vision sensor, which provided [details]. To allow convenient control of the vision sensor's movement, we attached it to the Pioneer P3DX robot model, which is included with V-Rep.

- The sequence of 2D images from the vision sensors~~ projection plane~~ is the input stream from which ground truth is extracted.
- Rather than use ~~some of~~ VREPs built in object detection algorithms, a camera plane projection matrix is used to determine objects that lie on the projection plane. This reduces computational complexity. and increases accuracy.
- The VREP simulation was run in synchronous mode with a timestep specified by the IT cortex model.
- Default timestep was set to 5ms.
- At initialization, the VREP simulation is queried for a list of all objects in the scene.
- Constant object attributes are also determined. This include the size of the object which is defined as twice the maximum length in the x, y or z dimension, the rotation symmetries (rotation period and mirror symmetric) in the x, y and z dimension and the object handles of all objects and their component parts.
- Object rotation symmetries are added manually under the custom data field. If no rotation symmetries are retrieved, defaults of rotation symmetry of one and not mirror symmetric are assumed.
- A population of model neurons that is selective for these objects is then generated, ~~Default population size is 100 neurons.~~ with a user-configurable size
- Each timestep the IT cortex model triggers the VREP simulation to run and queries it for ground truth using VREPs remote APIs.
- To get projection plane coordinates of each object (x_p, y_p), the coordinates of the objects center of mass in the reference frame of the vision center (x, y, z) is retrieved. Where the reference frame of the vision system is a coordinate system with its origin at the center of mass of the vision sensor.

$$\begin{bmatrix} x_{h_p} \\ y_{h_p} \\ z_{h_p} \\ w_{h_p} \end{bmatrix} = \begin{bmatrix} \frac{1}{ar*tan(\frac{\alpha}{2})} & 0 & 0 & 0 \\ 0 & \frac{1}{tan(\frac{\alpha}{2})} & 0 & 0 \\ 0 & 0 & \frac{-(z_n+z_f)}{(z_n-z_f)} & \frac{2z_nz_f}{(z_n-z_f)} \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{14}$$

where

$ar = \frac{screen\,width}{screen\,height}$ = aspect ratio of the projection plane,

α is the perspective angle of the vision sensor,

z_n is the near clipping distance of the vision sensor,

z_f is the far clipping distance of the vision sensor, the distance between the vision sensor and its projection plane,

w_{h_p} stores the value of z during the transformation.

ar, α, z_n, z_f are parameters of the vision sensor and were set to 1 (128x128 pixels), 90°, 0.02m, 2m respectively.

- projection plane coordinates are calculated using:

$$\begin{aligned} x_p &= \frac{x_{h_p}}{w_{h_p}} * ar * \frac{\pi}{2} \\ y_p &= \frac{y_{h_p}}{w_{h_p}} * \frac{\pi}{2} \end{aligned} \tag{15}$$

where

Multiplying $x_{h_p}$with ar normalizes the range of x_p to lie within ($-ar$, ar) and restores the aspect ratio of the projection plane.

y_{h_p} ranges between (-1, 1).

The x-axis of the projection plane is assumed to scan the whole visual space, 180°. Is this relevant?

x_p, y_p are converted to radians of eccentricity by multiple with $\frac{\pi}{2}$.

This is strange. You can't convert plane coordinates to angles over a large range of angles by multiplying by a constant.

- The projection plane size of an object in radians of eccentricity is given by

$$size_p = \frac{\pi * size}{2 * ar * z_e * tan\left(\frac{\alpha}{2}\right)}$$

where
$size$ is the real world size of the object,
z_e is the Euclidean distance of the object from the vision sensor,
and we have use the relationship $2 * ar = \pi$ to convert to radians of eccentricity.

- The orientation of an object is found by querying its Euler rotation angles (α, β, γ) with respect to the vision sensors frame.
- The vision sensor itself is rotated by $\beta = 90°$, $\gamma = 90°$ such that the vision sensors z-axis is along the real world x-axis (direction of motion of the IT cortex robot), y-axis is along the real world z-axis (vertical) and the x-axis is in the direction of the real world y-axis.
- ~~By getting the rotations with respect to the vision sensors reference frame, rotations of the vision sensor with respect to the real world reference frame can be ignored.~~
- *However, it is useful to know that rotations about the real world x-axis, map to rotations around the vertical y-axis of vision sensor. The IT cortex model defines rotation tuning profiles about the vision sensors y-axis only. Rotations of the objects around the other axis map onto rotations around the y and z axis.*

not relevant - this can be a code comment

you can simplify the explanation in the paper by giving the axes different names in the paper.

- A VREP child script is used to find the visibilities of the diagnostic and nondiagnostic parts of all objects. - This is a detail you have to expand on - get rid of some of the more obvious stuff above to make room.
- Each simulation step object handles for all objects and their component parts are passed to the child script.
- The vision sensor is configured to multiplex the object handle of each pixel with its value (in Vision sensor properties the Render mode field is set to OpenGL, color coded handles). *The latest version of VREP has render mode which can be set to ray traced image. This should return the object handle of each pixel. However, since this part is already working we do not use the ray traced render mode.*

- Too low-level - VREP implementation details belong in code comments

- Once all objects that lie in the projection plane are determined, their handles are sent to the child script. The child script decodes the object handle of each pixel and counts the number of visible pixels for each object.
- For each object, the script removes all other objects from the scene (by marking them as unrenderable), reacquires its projection image and counts the total number of object pixels.
- The ratio of visible to total pixels is used to determine visibilities levels of each object.
- Diagnostic parts of object are treated as a separate object thereby allowing the script to retrieve diagnostic parts visibilities as well.
- Once all ground truth is extracted it is fed into the IT Cortex model and the firing rate of each neuron is found.

more succinct

4 Results

4.1 Selectivity

4.2 Maximum Firing Rate

- *Include plot of mean firing rate vs selectivity. similar to Fig 5a of [9]. Log plot, shows firing rate decreases with selectivity defined as Kurtosis.*

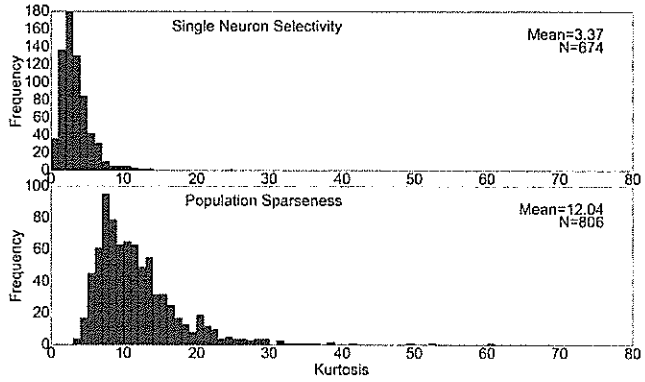


Figure 2: Neuron selectivity and population sparseness distributions. (A) Neuronal selectivity histogram. N=Number of model neurons. Average excess Kurtosis=3.37. (B) Population Sparseness histogram. N= Number of test stimuli. Average excess Kurtosis=12.04. We generated similar sized population of IT neurons and test stimuli set and successfully reproduced the result that population sparseness is greater than neuron selectivity seen in [9] with our adjusted parameter distributions.

- it will be best to reproduce the source figures with permission where possible.

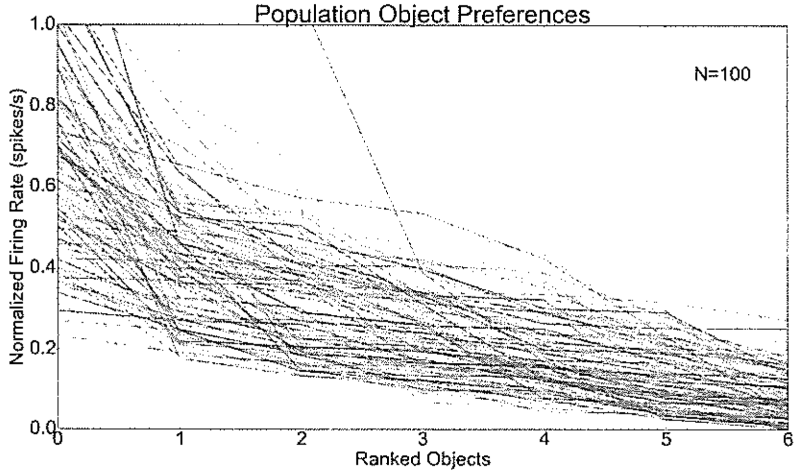
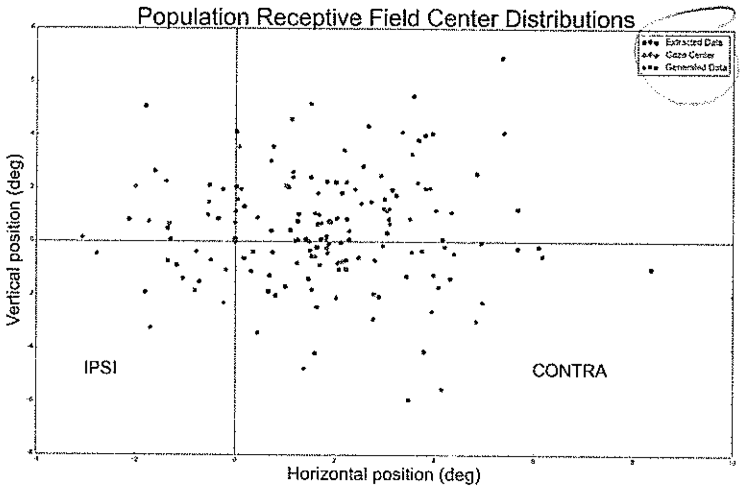


Figure 3: Plot of ranked normalized object selectivity of neurons in a generated population. Objects with the highest selectivities are plotted first. *Similar to Figure 2 in [18]. Does this figure add any value?*

I think it's good if we put it beside the original



Too small.

Figure 4: Distribution of Receptive Field centers around the center of gaze. Extracted data from [14].

4.3 Position

- Figure 4 plots receptive field centers across a generated population with receptive field centers found in [14].

4.4 Size

4.5 Orientation

4.6 Occlusion

4.7 Clutter

4.8 Dynamic

5 Discussion

- begin the discussion with a short recap of what you did, how well it worked (pretty well in that you reasonably fit all the data you tried to fit)

off topic

- Comparison with other methods of
  - [7] representation in visual system can be classified into three categories:
    - Tuning Curve measurements: Isolate a single stimulus dimension. Measure responses to stimuli that vary along that dimension.
    - Multivariate pattern classification: Again isolate a single stimulus, vary stimulus along the dimension, but look at multiple neuron responses and train a classifier to predict the dimension of interest from measured responses.
    - Receptive field Estimation measure responses to a large number of stimuli that vary along multiple dimensions. Develop a single model that models responses over multiple dimensions. Example Gabor filter, looks at spatial freqency, contrast/amplitude and orientation.
  - [7], problems with tuning curve measurement results
    - difficult to isolate stimuli dimensions.

- explain why the work is useful:
  - input to models of deeper areas; sensitivity analysis of IT params for deeper functio
  - point of reference for statistical comparison with more abstract IT models like upper layers of convnets
  - to highlight unknowns in the literature (give examples of things you had to

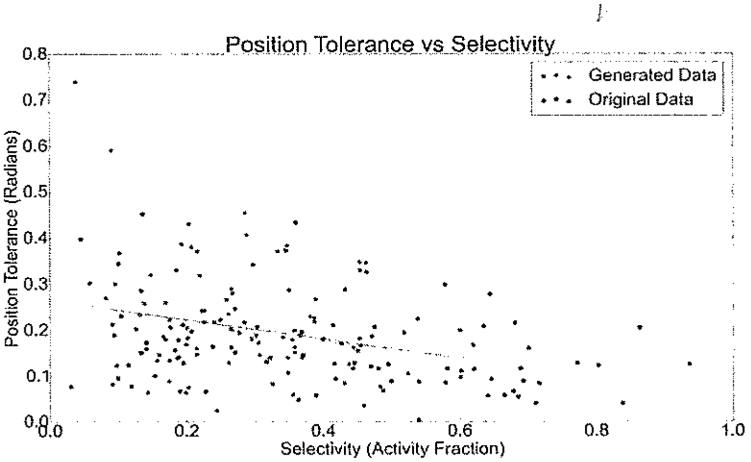


Figure 5: Postion Tolerance verses Selectivity. [18] found a negative correlation between activity fraction selectivity and position tolerance of a neuron. As the neuron become more selective its position tolerance decreased - IT neurons that respond to objects are not grandmother cells. In this respect, IT neurons also differ from top level nodes of convolutional neuronal networks which have been compared with IT neurons but are designed to be position invariant. *Explain why the there are no activity fractions higher that 0.6 in the generated but have been seen in recordings of IT neurons. Consequence of using gamma distribution model to generate selectivity profiles? Increases the population size or the number of test stimuli still does not generate activity fractions greater than 0.6.*

Not really. They're very slightly posit. invariant in a convnet. due to poolin

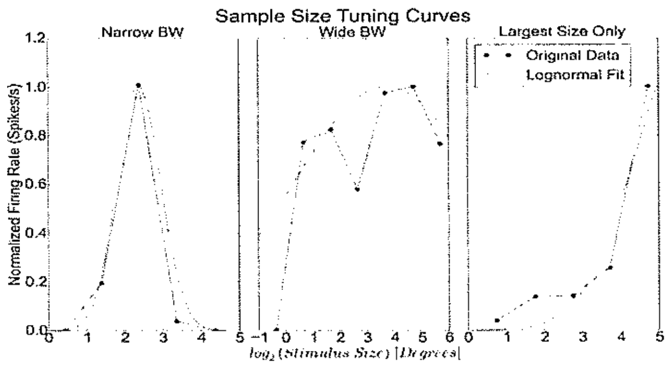


Figure 4. Types of Size tuning profiles found in IT cortex and lognormal fits. (A) Example Neuron with narrow tuning (size bandwidth < 2 octaves). (B) Example neuron with broad tuning (size bandwidth > 5 octaves) (Majority). (c) Example Neuron that only showed significant response to the largest stimulus size tested. Original data from Ito et. al - 1995.

Figure 6: Types of size tuning profiles found in IT cortex and lognormal fits. (A) Example neuron with narrow tuning (size bandwidth < 2 octaves). (B) Example neuron with broad tuning (size bandwidth > 5 octaves). (c) Example neuron that only showed significant response to the largest stimulus size tested. Original data from [6]. . *Fix figure why is there a caption within the caption.*

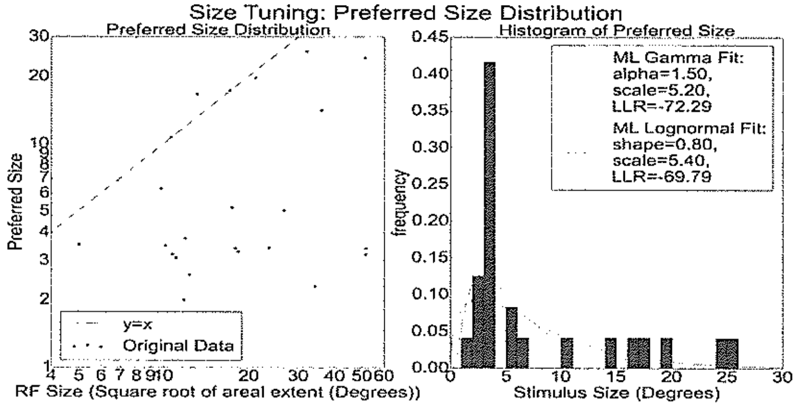


Figure 5. Distribution and histogram of preferred stimulus size of IT neurons. Original data from Ito et. al - 1995.

Figure 7: Distribution and histogram of preferred stimulus size of IT neurons. Original data from Ito et. al - 1995. *This figure isn't very useful only data fitting. Should include generated data distribution as a comparison. Also two different fits are shown gamma and lognormal. This is not explained in the text and should be added. How else to improve? Fix figure why is there a caption within the caption.*

These changes sound good - also quantify this if you are comparing two

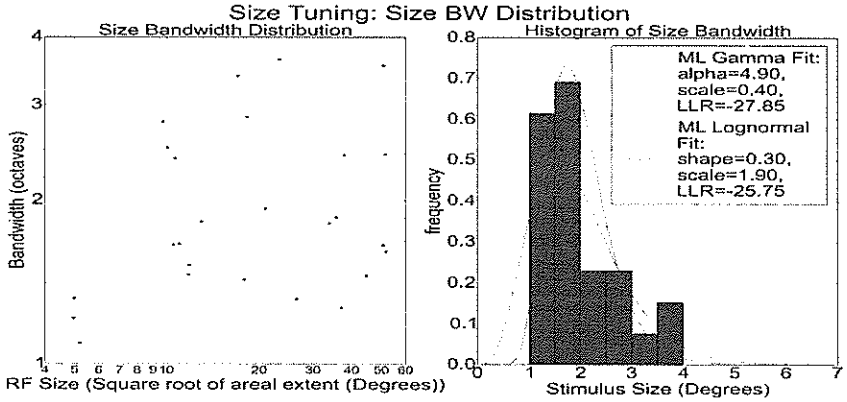
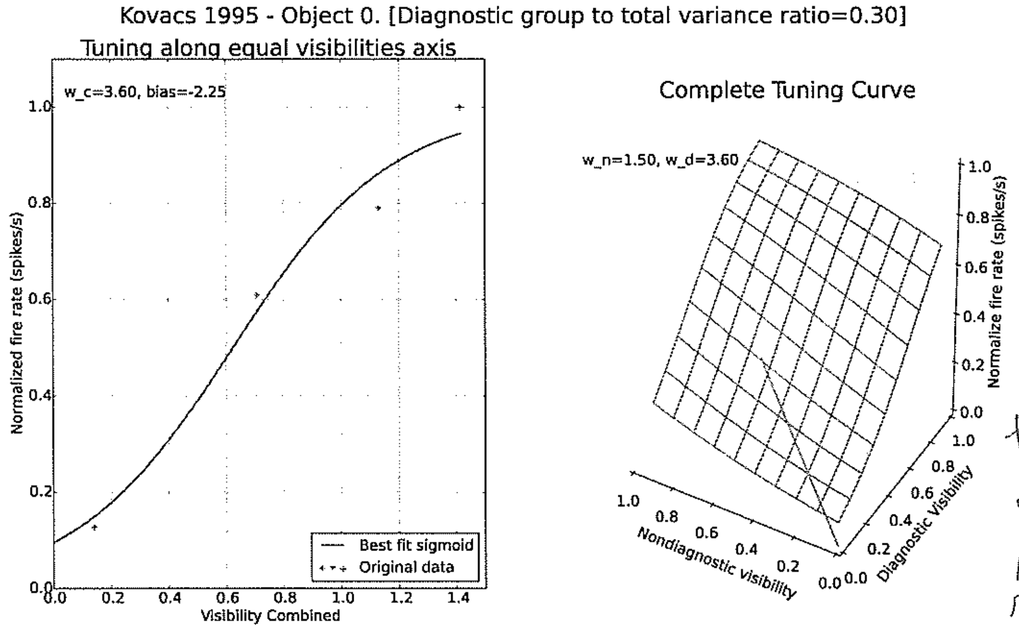


Figure 6. Distribution and histogram of size bandwidths of IT neurons. Original data from Ito et. al - 1995.

Figure 8: Distribution and histogram of size bandwidths of IT neurons. Original data from Ito et. al - 1995. *This figure isn't very useful, only data fitting. Should include generated data distribution as a comparison. Also two different fits are shown gamma and lognormal. This is not explained in the text and should be added. How else to improve? Fix figure why is there a caption within the caption.*

Need to generate figure.

Figure 9: Rotation tuning profiles for objects with various rotation symmetries. In [6], three types of rotation tuning profiles were found: single peak, bimodal and invariant (A) No mirror symmetry and no symmetry period (single peak - most objects), (B) Mirror symmetry but no symmetry period (bimodal - faces, cars), (C) Mirror symmetry and symmetric period of 4 (Cube (multiple peaks)), (D) Mirror Symmetry and symmetric period of 360° (view invariant).



instead of the line along c, 3D line plot along surface.

Figure 10: Sample Occlusion Tuning Profile. We fit our occlusion model to the normalized responses of object ranked 1 in the moving occluder data tuning curve of [8] and found the best fit w_c and b. Given a diagnostic parts preference, R, we used nonlinear optimization to determine a set of w_d and w_{nd} that can generate R. See methods section for details. As noted in[13], both diagnostic and nondiagnostic parts are needed to get the unoccluded firing rate.

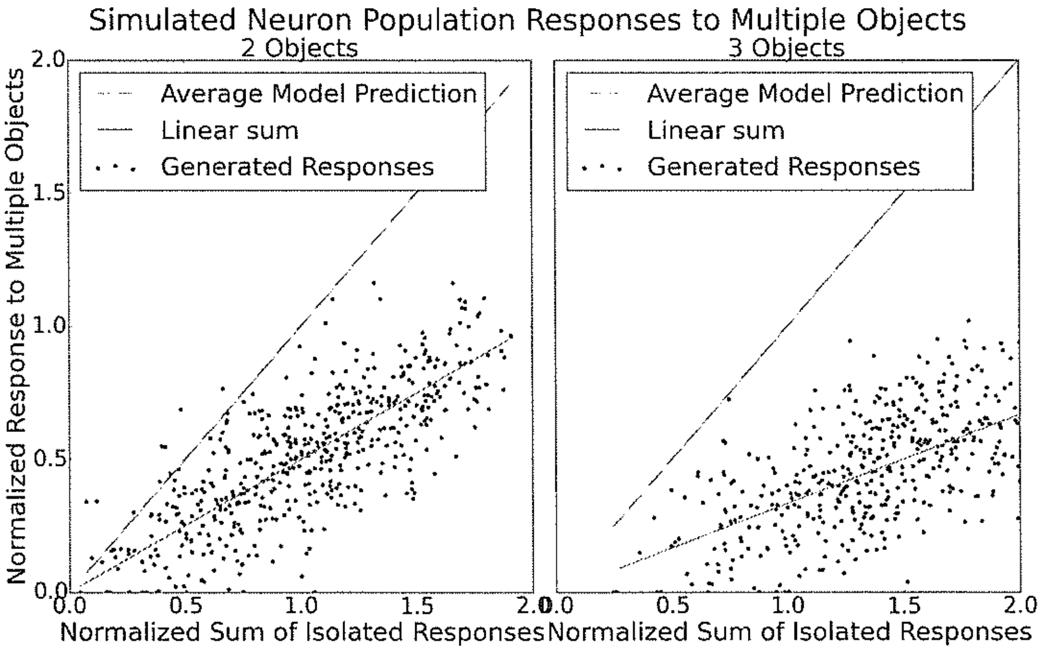


Figure 11: *Fill me in*

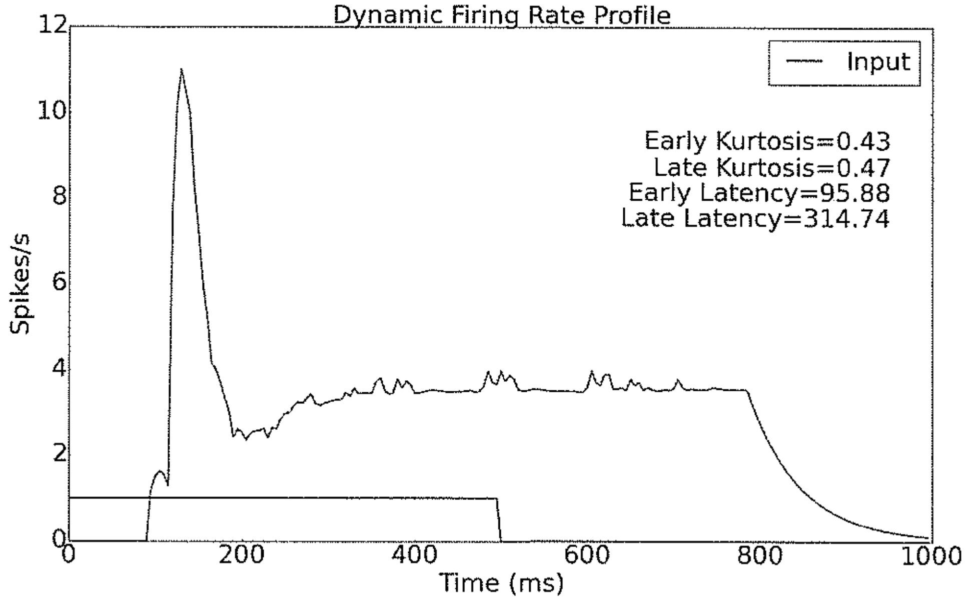


Figure 12: *Fill me in. Displayed Kurtosis values are incorrect. These are using the normalized firing to calculating Kurtosis. Should use the normalized firing rates times the max firing rate. Redo.*

    - * isolated stimulus may not be the actual dimension that IT neurons are responded to, it may be correlated to the actual stimulus dimension.
    - * difficult to generalize results

mention this as a limitation

  - [7] issues with developing a single receptive field model for inferior temporal neurons
    - * high dimensional space
    - * IT responds to complex stimuli which cannot be uniquely quantified.

good

- How many tuning curves and object attributes to consider before the IT cortex model can approach the real IT Cortex
  - [10] estimated the intrinsic dimensionality of the IT cortex to be 93±11SD.
  - Most papers use stimuli sample sizes and record from neurons smaller than this dimensionally. *so what?*
  - Difficult to come up with a single receptive field model (with multiple stimuli dimensions) as IT neurons deal with high dimensional stimuli only a few of which have been identified and even fewer have been completely understood.

References

[1] Rohatgi Ankit. Webplotdigitizer, 2015.

[2] James J DiCarlo, Davide Zoccolan, and Nicole C Rust. How does the brain solve visual object recognition? *Neuron*, 73(3):415–434, 2012.

[3] M. Freese E. Rohmer, S. P. N. Singh. V-rep: a versatile and scalable robot simulation framework. In *Proc. of The International Conference on Intelligent Robots and Systems (IROS)*, 2013.

[4] Ichiro Fujita, Keiji Tanaka, Minami Ito, and Kang Cheng. Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, 360(6402):343–346, 1992.

[5] Chou P Hung, Gabriel Kreiman, Tomaso Poggio, and James J DiCarlo. Fast readout of object identity from macaque inferior temporal cortex. *Science*, 310(5749):863–866, 2005.

[6] Minami Ito, Hiroshi Tamura, Ichiro Fujita, and Keiji Tanaka. Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of neurophysiology*, 73(1):218–226, 1995.

[7] Kendrick N Kay. Understanding visual representation by developing receptive-field models. *Visual Population Codes: Towards a Common Multivariate Framework for Cell Recording and Functional Imaging*, pages 133–162, 2011.

[8] GY Kovács, Rufin Vogels, and Guy A Orban. Selectivity of macaque inferior temporal neurons for partially occluded shapes. *The Journal of Neuroscience*, 15(3):1984–1997, 1995.

[9] Sidney R Lehky, Roozbeh Kiani, Hossein Esteky, and Keiji Tanaka. Statistics of visual responses in primate inferotemporal cortex to object stimuli. *Journal of neurophysiology*, 106(3):1097–117, sep 2011.

[10] Sidney R Lehky, Roozbeh Kiani, Hossein Esteky, and Keiji Tanaka. Dimensionality of object representations in monkey inferotemporal cortex. *Neural computation*, 2014.

[11] Nuo Li, David D Cox, Davide Zoccolan, and James J DiCarlo. What response properties do individual neurons need to underlie position and clutter invariant object recognition? *Journal of neurophysiology*, 102(1):360–376, 2009.

[12] Nikos K Logothetis, Jon Pauls, and Tomaso Poggio. Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, 5(5):552–563, 1995.

[13] Kristina J Nielsen, Nikos K Logothetis, and Gregor Rainer. Dissociation between local field potentials and spiking activity in macaque inferior temporal cortex reveals diagnosticity-based encoding of complex objects. *The Journal of neuroscience*, 26(38):9639–9645, 2006.

[14] Hans Op De Beeck and Rufin Vogels. Spatial sensitivity of macaque inferior temporal neurons. *Journal of Comparative Neurology*, 426(4):505–518, 2000.

[15] Edmund T Rolls, Nicholas C Aggelopoulos, and Fashan Zheng. The receptive fields of inferior temporal cortex neurons in natural scenes. *The Journal of Neuroscience*, 23(1):339–348, 2003.

[16] Keiji Tanaka. Inferotemporal cortex and object vision. *Annual review of neuroscience*, 19(1):109–139, 1996.

[17] Davide Zoccolan, David D Cox, and James J DiCarlo. Multiple object response normalization in monkey inferotemporal cortex. *The Journal of Neuroscience*, 25(36):8150–8164, 2005.

[18] Davide Zoccolan, Minjoon Kouh, Tomaso Poggio, and James J DiCarlo. Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *The Journal of Neuroscience*, 27(45):12292–12307, 2007.