massachusetts institute of technology — artificial intelligence laboratory

Categorization in IT and PFC: Model and Experiments

Ulf Knoblich, David J. Freedman and Maximilian Riesenhuber

AI Memo 2002-007 April 2002
CBCL Memo 216

Abstract

In a recent experiment, Freedman *et al.* recorded from inferotemporal (IT) and prefrontal cortices (PFC) of monkeys performing a "cat/dog" categorization task ([3] and Freedman, Riesenhuber, Poggio, Miller, *Soc. Neurosci. Abs.*). In this paper we analyze the tuning properties of view-tuned units in our HMAX model of object recognition in cortex [7, 8] using the same paradigm and stimuli as in the experiment. We then compare the simulation results to the monkey inferotemporal neuron population data. We find that view-tuned model IT units that were trained without any explicit category information can show category-related tuning as observed in the experiment. This suggests that the tuning properties of experimental IT neurons might primarily be shaped by bottom-up stimulus-space statistics, with little influence of top-down task-specific information. The population of experimental PFC neurons, on the other hand, shows tuning properties that cannot be explained just by stimulus tuning. These analyses are compatible with a model of object recognition in cortex [10] in which a population of shape-tuned neurons provides a general basis for neurons tuned to different recognition tasks.

This report describes research done within the Center for Biological & Computational Learning in the Department of Brain & Cognitive Sciences and in the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology.

This research was sponsored by grants from: Office of Naval Research (DARPA) under contract No. N00014-00-1-0907, National Science Foundation (ITR) under contract No. IIS-0085836, National Science Foundation (KDI) under contract No. DMS-9872936, and National Science Foundation under contract No. IIS-9800032.

Additional support was provided by: AT&T, Central Research Institute of Electric Power Industry, Center for e-Business (MIT), Eastman Kodak Company, DaimlerChrysler AG, Compaq, Honda R&D Co., Ltd., ITRI, Komatsu Ltd., Merrill-Lynch, Mitsubishi Corporation, NEC Fund, Nippon Telegraph & Telephone, Oxygen, Siemens Corporate Research, Inc., Sumitomo Metal Industries, Toyota Motor Corporation, WatchVision Co., Ltd., and The Whitaker Foundation. M.R. is supported by a McDonnell-Pew Award in Cognitive Neuroscience.

1 Introduction

In [10], Riesenhuber and Poggio proposed a model of object recognition in cortex in which a general representation of objects in inferotemporal cortex (IT) provides the basis for different recognition tasks — such as identification and categorization — with task-related units located further downstream, *e.g.,* in prefrontal cortex (PFC). Freedman and Miller recently performed physiology experiments providing experimental population data for both PFC and IT of a monkey trained on a "cat/dog" categorization task ([2, 3] and Freedman, Riesenhuber, Poggio, Miller, *Soc. Neurosci. Abs.*, 2001). In this paper, using the same stimuli as in the experiment, we analyze the properties of view-tuned units in our model, trained without any explicit category information, and compare them to the tuning properties of experimental IT and PFC neurons.

2 Methods

2.1 The HMAX model

We used the hierarchical object recognition system of Riesenhuber & Poggio [7, 8], shown schematically in Fig. 1. It consists of a hierarchy of layers with linear units performing template matching, and non-linear units performing a "MAX" operation. This MAX operation, selecting the maximum of a cell's inputs and using it to drive the cell, is key to achieving invariance to translation, by pooling over afferents tuned to different positions, and scale, by pooling over afferents tuned to different scales. The template matching operation, on the other hand, increases feature specifity. A cascade of these two operations leads to C2 units (roughly corresponding to V4/PIT neurons), which are tuned to complex features invariant to changes in position and scale. The outputs of these units (or a subset thereof) are used as inputs to the view-tuned units (corresponding to view-tuned neurons in IT [5, 8]), which in turn can provide input to units trained on various recognition tasks, for instance cat/dog categorization (for the appropriate simulations, see [9]).

2.2 Stimulus space

The stimulus space is spanned by six prototype objects, three "cats" and three "dogs" (cf. Fig. 2). Our morphing software [11] allows us to generate 3D objects that are arbitrary combinations of the six prototypes. Each object is defined by a six-dimensional morph vector, with the value in each dimension corresponding to the relative proportion of one of the prototypes present in the object. The component sum of each object was constrained to be equal to one. An object was labeled a "cat" or "dog" depending on whether the sum over the "cat" prototypes in its morph vector was greater or smaller than those over the "dog" prototypes, resp. The

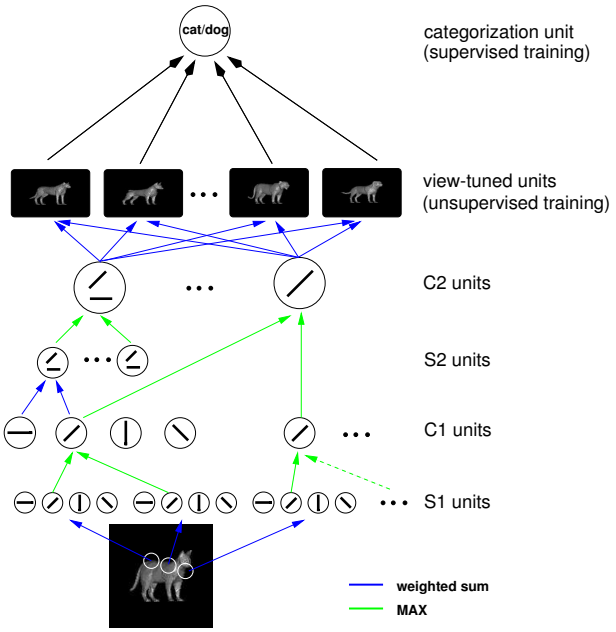

Figure 1: Scheme of the HMAX model. Feature specifity and invariance to translation and scale are gradually built up by a hierarchy of "S" and "C" layers [4], resp. The C2 layer, consisting of units tuned to complex features invariant to changes in position and scale, feeds directly into the view-tuned units, which in turn can provide input to recognition task-specific units, such as a cat/dog categorization unit, as shown (see [9]).

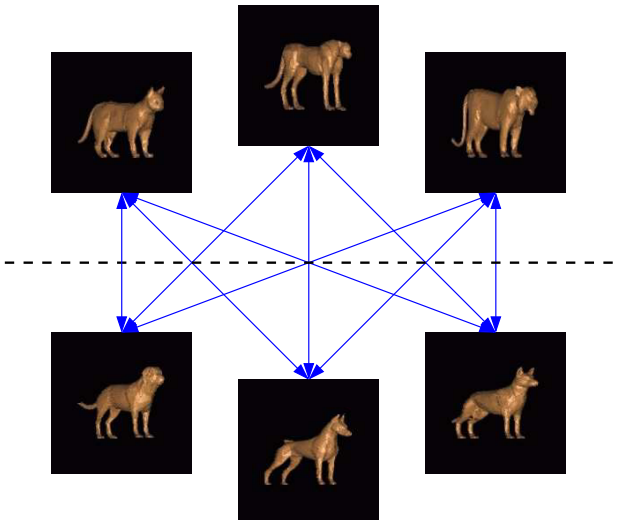

Figure 2: Illustration of the cat/dog stimulus space. The morph space is spanned by the pictures of three cats shown on top ("house cat", "Cheetah" and "Tiger") and the three dogs below ("house dog", "Doberman" and "German Shepherd"). All prototypes have been normalized with respect to viewing angle, lighting parameters, size and color.

class boundary was defined by the set of objects having morph vectors with equal cat and dog component sums. The lines in Fig. 2 show the nine possible morph lines between two prototypes, one of each class, as used in the test set (see below).

Training set. The training set (a subset of the stimuli used to train the monkeys in [2, 3]) consisted of 144 randomly selected morphed animal stimuli not restricted to these morph lines [2], but chosen at random from the cat/dog morph space, excluding "cats" ("dogs") with a "dog" ("cat") component greater than 40% (as in the experiment).

Test set. The testing set used to determine an experimental neuron's or model unit's category tuning consisted of the nine lines through morph space connecting one prototype of each class. Each morph line was subdivided into 10 intervals, with the exclusion of the stimuli at the mid-points (which would lie right on the class boundary, with an undefined label), yielding a total of 78 stimuli.

2.3 Learning a class representation

One view-tuned unit (VTU), connected to all or a subset of C2 units, was allocated for each training stimulus, yielding 144 view-tuned units.* The two parameters affecting the tuning characteristics of the VTUs are the number of afferent C2 units a (sorted by decreasing strength [7]) and the Gaussian tuning width σ. Experiments were run using 8, 32, 128 and 256 afferents to each VTU and σ values of 0.1, 0.2, 0.4, 0.8 and 1.6, respectively. For the sake of clarity we will present only four of those 20 combinations ($(a = 32, \sigma = 0.1)$; $(a = 32, \sigma = 0.2)$; $(a = 256, \sigma = 0.1)$ and $(a = 256, \sigma = 0.2)$). Using 8 afferents produced units whose tuning was too unspecific, while σ values above 0.2 yielded unrealistically broad tuning.

2.4 Evaluating category tuning

We use three measures to characterize the category-related behavior of experimental neurons and model units: the between-within index (BWI), the class coverage index (CCI) and the receiver operating characteristics (ROC).

BWI The *between-within index* (BWI) [2, 3] is a measure for tuning at the class boundary relative to the class interior. Considering the response of a unit to stimuli along one morph line, the response difference between two adjacent stimuli can be calculated. As there is no stimulus directly on the class boundary, we use 20% steps for calculating the response differences. Let btw be the mean response difference *between* the two categories (*i. e.*, between morph index 0.4 and 0.6) and

*Results were similar for 32 VTUs obtained from 144 stimuli through k-means clustering [9].

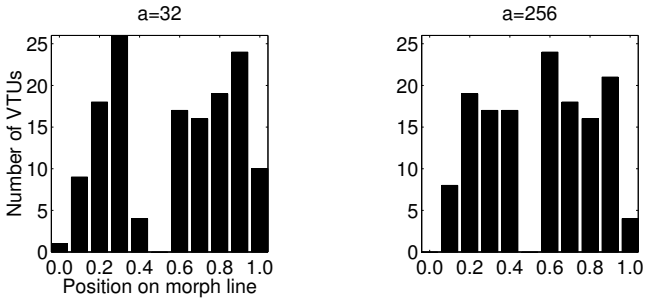

Figure 3: Number of view-tuned units tuned to stimuli at certain morph indices for different numbers of afferents. The morph index is the percentage of the dog prototype in the stimulus, *e. g.*, morph index 0.4 corresponds to a morphed stimulus which is 40% dog and 60% cat.

wi the mean response difference *within* the categories. Then the between-within index is

$$\text{BWI} = \frac{btw - wi}{btw + wi}. \quad (1)$$

Thus, the range of BWI values is -1 to $+1$. For a BWI of zero the unit shows on average no different behavior at the boundary compared to the class interiors. Positive BWI values indicate a significant response drop across the border (*e. g.*, for units differentiating between classes) whereas negative values are characteristic for units which show response variance within the classes but not across the boundary.

CCI The *class coverage index* (CCI) [2] is the proportion of stimuli in the unit's preferred category that evoke responses higher than the maximum response to stimuli from the other category. Possible values range from $\frac{1}{39}$, meaning out of the 39 stimuli in the class only the maximum itself evokes a higher response than the maximum in the other class, to 1 for full class coverage, *i. e.*, perfect separability.†

ROC The *receiver operating characteristics* (ROC) curve [6] shows the categorization performance of a unit in terms of correctly categorized preferred-class stimuli (hits) *vs.* miscategorized stimuli from the other class (false alarms). The area under the ROC curve is a measure of the quality of categorization. A value of 0.5 corresponds to chance performance, 1 means perfect separability, *i. e.*, perfect categorization performance. ROC values were obtained by fixing the activation threshold and counting hits and false alarms using this threshold. Activation values for all stimuli were used as thresholds to obtain ROC curves as detailed as possible. The ROC

†When comparing model units and experimental neurons, CCI values were calculated using the 42 stimuli used in the experiment (see section 4), so the minimum CCI value was $\frac{1}{21}$.

area values were computed using trapezoidal numerical integration of the ROC curve.

3 Results

3.1 VTU positions

We defined the position of a VTU as the position of the stimulus along the nine morph lines that maximizes the VTU's response. This approach yields an asymmetric distribution of VTUs over the stimulus space (Fig. 3). As the training set consisted of 72 cats and 72 dogs, this asymmetry suggests that some morphed "cats" look similar to "dogs" in the space of C2 activations.

3.2 Shape tuning

The response function over the stimulus space of a view-tuned unit is a Gaussian centered at the unit's preferred stimulus, dropping exponentially fast (depending on the unit's standard deviation σ) in all directions. Thus, the VTU will respond to every stimulus presented, however there will be a significant response to only some of those stimuli.

Fig. 4 shows the response to the stimuli along the nine morph lines of a VTU tuned close to the 80% cat stimulus on the fifth line (connecting Cheetah and Doberman). For small σ, the tuning is much tighter around this maximum, showing only little response to stimuli involving other cat prototypes. With 32 afferents and a σ of 0.1 this unit's behavior could be described as categorizing cats *vs.* dogs along the morph lines involving cat prototype 2. With 256 afferents the tuning is even tighter because the stimulus is specified in a 256-dimensional space instead of a 32-dimensional one. Because the unit is tuned to a randomly generated stimulus not necessarily lying on a morph line, there is hardly any response for $a = 256$ and $\sigma = 0.1$.

3.3 Category tuning

3.3.1 BWI

The between-within index is a measure for response changes at the class boundary *vs.* the class interiors. For a given parameter set, all units share the same response function shape, only varying in the location of their preferred stimulus in morph space. Thus, the major response decay of prototype-tuned units will be within their class, showing no observable response to stimuli near the border or in the other class, yielding negative BWI values. Border-tuned units show a substantial drop in response over the boundary with a much lower mean difference within class resulting in a positive between-within index. As shown in Fig. 5, for units with preferred stimuli at different positions along each morph line, units tuned with 256 afferents show exactly this behavior. However, when using only 32 afferent VTUs the values tend to be closer to zero. This is due to

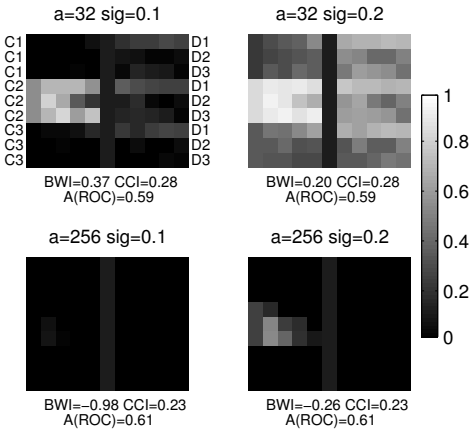

Figure 4: Grayscale plot of a VTU response along the nine morph lines. Each horizontal line represents one cross-boundary morph line. The vertical middle line is inserted to visually separate cat (left) and dog (right)stimuli. Cat prototypes (C1, C2, C3) are plotted to the left of the boundary, dog prototypes (D1, D2, D3) to the right. The columns in between correspond to morphed stimuli in 10% steps. A color scale indicates the response level.

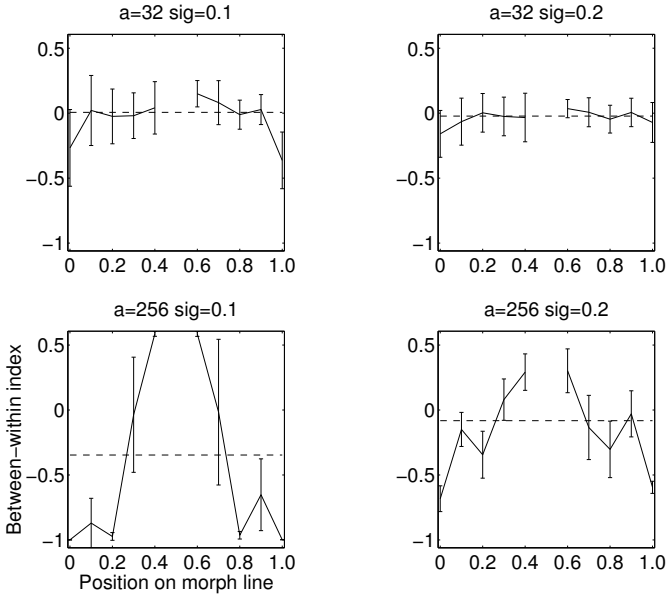

Figure 5: Mean between-within index (BWI) of view-tuned units with preferred stimuli spaced along the morph lines. Error bars show standard deviation across the nine morph lines.

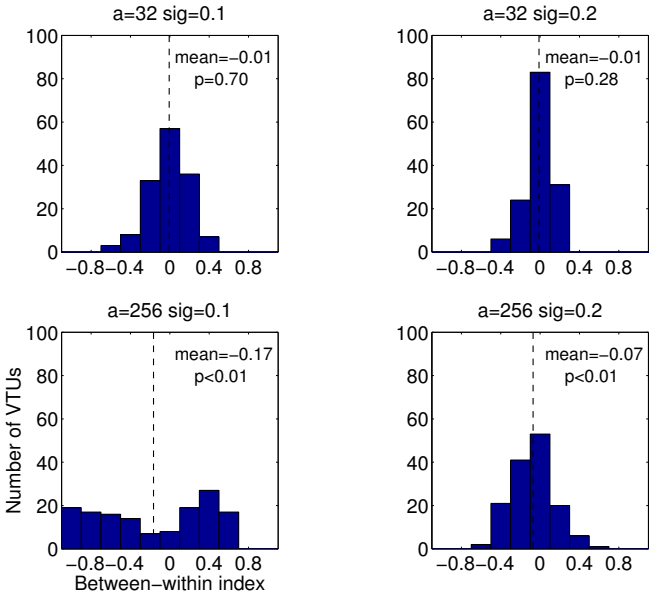

Figure 6: Histograms of between-within index (BWI) values for the 144 model VTUs. The dashed line indicates the mean between-within index over all view-tuned units.

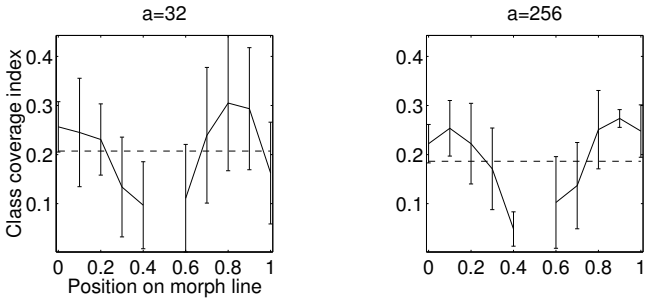

Figure 7: Mean class coverage index (CCI) of view-tuned units with preferred stimuli spaced along the morph lines. Error bars show standard deviation across the nine morph lines.

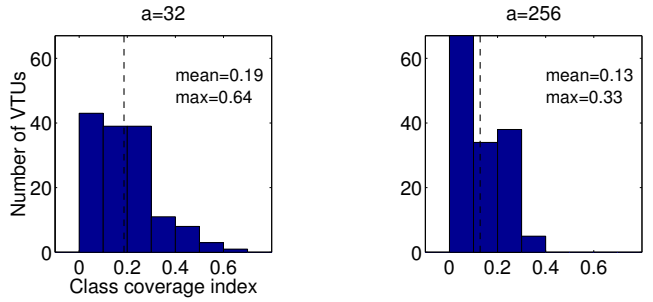

Figure 8: Histograms of class coverage index (CCI) values. The dashed line indicates the mean class coverage index over all view-tuned units.

the less precise tuning of those VTUs (cf. Fig 4), yielding smaller absolute values of the BWI.

The histograms of BWI values (Fig. 6) for the 144 units reflect this fact. The broader the VTU tuning gets with decreasing a and increasing σ, the tighter the distribution of BWI values is centered around zero. For 256 afferents, there is a significant shift of the whole distribution towards negative values ($p < 0.01$).

3.3.2 CCI

The class coverage index does not depend on σ because changing σ will change the width of the Gaussian but not its shape or position. Clearly, the CCI of a VTU is dependent on the position of its preferred stimulus in morph space. Units tuned to stimuli near the class boundary will have lower CCI values because the response level to stimuli on the other side of the border will be quite high. The class coverage index for units tuned to stimuli near the center of a class (*e. g.*, at morph line positions 0.2 and 0.8) will be higher, as the maximum response to an other-class stimulus will be lower because of the bigger distance from the center of tuning to the class boundary. Units tuned to the prototype stimuli will have smaller CCI values again due to the visual dissimilarity of the prototypes (cf. Fig. 7). As can be seen from Fig. 2, the visual appearance of prototypes of the same class is quite different. Thus, *e. g.*, the distance in the space of C2 activations from the Tiger prototype to the morphed stimulus which is only 40% Tiger and 60% Doberman is smaller than the distance from the Tiger prototype to the Cheetah prototype. As indicated by the error bars, there is a wide variance for the class coverage index at one position on a morph line.

As can be seen in Fig. 8, the VTU's CCI values are shifted towards zero when increasing the number of afferents, since the more specific tuning will emphasize the dissimilarity of prototypes of the same class. For $a = 32$, there are some units with CCI values of 0.4 and above. This means those units show a response behavior similar to categorization in certain parts of stimulus space. Fig. 9 shows the response of a single VTU along the nine morph lines. With the unit's category threshold at the indicated position the categorization performance is 85%.

3.3.3 ROC

The CCI value corresponds (up to a factor) to the number of stimuli that evoke responses higher than the maximum other-class stimulus. Thus, this value is equivalent to the number of correctly categorized stimuli with no false alarms (*i. e.*, no other-class stimulus being miscategorized), which is the initial point of the ROC curve. Fig. 11 shows the distribution of A_{ROC} values for different numbers of afferents (as changing σ only affects the response magnitude to individual stimuli but does not change their ranking, A_{ROC} is independent of σ). For 256 afferents, about half of the units

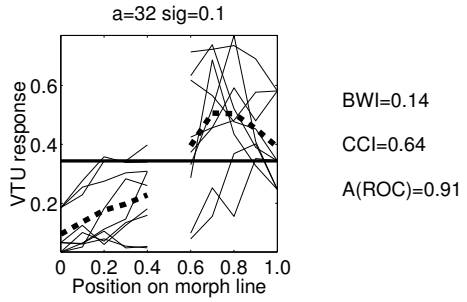

Figure 9: Response of a view-tuned unit with maximum response to an 80% dog stimulus (one of 144 VTUs, 32 afferents to each VTU, $\sigma = 0.1$) along the nine morph lines. The dashed line shows the average over all morph lines. The solid horizontal line shows a possible class boundary yielding best categorization performance.

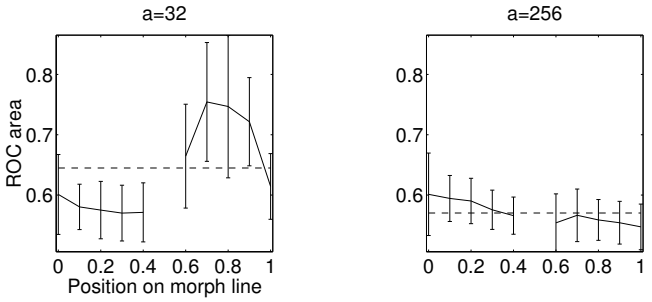

Figure 10: A_{ROC} values of view-tuned units with preferred stimuli spaced along the morph lines. Error bars show standard deviation across the nine morph lines.

have values over 0.6 with a maximum of 0.74. For 32 afferents about 15% of the VTUs have an A_{ROC} value of more than 0.8 up to 0.94. This clearly shows that there is a substantial number of VTUs able to categorize with a remarkable performance, without the benefit of any category information during training.

4 Comparison of model and experiment

We compared the tuning properties of model units to those of the IT and PFC neurons recorded from by Freedman [2, 3] from two monkeys performing the cat/dog categorization task.[‡] In the following, we restrict our analysis to the neurons that showed stimulus selectivity by an ANOVA ($p < 0.01$), over the 42 stimuli along the nine morph lines used in the experiment (in the experiment, stimuli were located at positions 0, 0.2, 0.4, 0.6, 0.8, and 1 along each morph line). Thus, we only analyzed those neurons that responded signif-

[‡]The monkeys had to perform a delayed match-to-category task. The first stimulus was shown for 600ms, followed by a 1s delay and the second, test, stimulus. See [2, 3] for details.

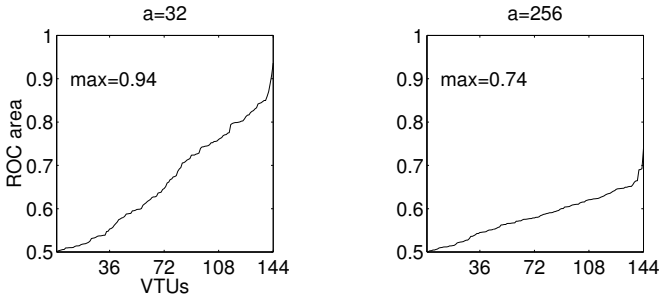

Figure 11: A_{ROC} values of the view-tuned units sorted in ascending order.

icantly differently to at least one of the stimuli.[§]

In particular, we analyzed a total of 116 stimulus-selective IT neurons during the "sample" period (100ms to 900ms after stimulus onset). Only a small number of IT neurons responded selectively during the delay period. For the PFC data, there were 67 stimulus-selective neurons during the sample period, and 32 stimulus-selective neurons during the immediately following "delay" period (300 to 1100 ms after stimulus offset, during which the monkey had to keep the category membership of the previously presented sample stimulus in mind, to compare it to a subsequently (at 1000 ms after stimulus offset) presented test stimulus [3].

Figs. 13 through 15 show the BWI, CCI, and A_{ROC} distributions for the IT neurons (during the sample period — IT neurons tended to show much less delay activity than the PFC neurons), and the PFC neurons (during sample and delay periods, resp.).[¶]

4.1 IT

Comparing the view-tuned model unit data to the experimental IT data (Fig. 16 and Fig. 13), we observe a very good agreement of the BWI distributions of model units and IT neurons: Both are centered around zero and show a mean not significantly different from 0. Further, the ROC plots show very similar means, and — even more importantly — identical maxima (0.94). This shows that high ROC values can be obtained without any explicit category information, and moreover that the range of ROC values of experimental IT neurons are well compatible with those of view-tuned model units. There do appear to be some differences in the distribution of ROC values, with the experimental distribution having proportionally fewer neurons with intermediate ROC values.

[§]Extending the analysis to include all *responsive* neurons (relative to baseline, $p < 0.01$) added mainly untuned neurons with CCIs close to 0, and A_{ROC} values close to 0.5.

[¶]For comparison with the model, the indices and ROC curves were calculated using a neuron's averaged firing rate (over at least 10 stimulus presentations) to each stimulus.

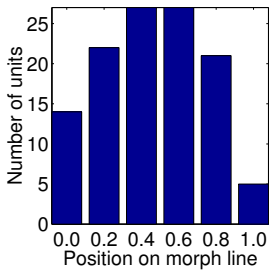

Figure 12: Distribution of preferred stimuli (morph indices) for experimental IT neurons.

Differences in the CCI distributions appear to be more substantial. However, the highest CCI in the model is greater than that of the experimental IT neurons, showing that model units can show similar degrees of category tuning as the experimental neurons.

4.1.1 Noise

What could be the source of the differences between the tuning properties of model units and experimental neurons? One factor is the deterministic response of model units in contrast to the noisy responses of experimental neurons which show trial-to-trial variations even for the same stimulus. Such random fluctuations in a neuron's firing rate can have a strong impact on the CCI value of neurons with preferred stimuli near the class boundary, where stimuli belonging to different classes produce similar responses, pointing to a possible explanation for the high number of neurons in the experiment with low CCI values.

Indeed, adding independent Gaussian noise to the responses of model units produces only modest shifts in the BWI and ROC distributions, but leads to a CCI distribution that is dominated by units with low CCI values, as in the experiment (Fig. 17). In the ROC value distribution, the proportion of units with intermediate ROC values decreased, producing a more "convex" shape as in the experiment. In general, the agreement with the experimental distribution is excellent, BWI and ROC distributions are not statistically significantly different ($p \geq 0.2$, Wilcoxon rank sum test), and the CCI distribution is only marginally different ($p = 0.06$).

4.1.2 Resampling

Another factor that might affect the population tuning properties is the distribution of preferred stimuli. In fact, calculating the distribution of preferred stimuli of the experimental IT neurons reveals a difference between experimental and model populations: As Fig. 12 shows, almost half of all experimental neurons have preferred stimuli at the class boundary, whereas the model units have a distribution that contains more neurons tuned to morph line centers (Fig. 3).

This difference could either be the signature of task-dependent influences on IT learning, or it could be due to statistics of the stimulus ensemble, as the later stages of the monkeys' training focussed on stimuli close to the boundary (which where most difficult for the monkeys to learn). It will be interesting to examine this question more closely in future studies where stimulus exposure is better controlled.

We investigated the effect of the distribution of preferred stimuli on the population tuning properties by *resampling* from the population of 144 model units to obtain a population with a distribution of preferred stimuli as in Fig. 12. Afterwards, a population of 116 units, the size of the experimental population, was drawn from those noisy units such that the distribution of units over the morph indices fit the experimental population. This procedure was repeated for a total of 100 trials and the results were averaged to obtain the correct number of values.

Fig. 18 shows the population tuning properties of the resampled model distribution (with a distribution of preferred stimuli as in Fig. 12), chosen from the deterministic model units of Fig. 16. Interestingly, distributions are not very different from the non-resampled case, in line with the results in Figs. 5, 7, and 10 that show only modest changes in the index values for units with preferred stimuli at the border compared to directly adjacent positions.

We investigated the combined effect of noise and distribution of preferred stimuli on population tuning properties by adding independent Gaussian noise with amplitude n to the responses of model units, and resampling from the population of 144 model units to obtain a population with a distribution of preferred stimuli as in Fig. 12. In particular the morph indices of the noisy model units were determined *after* adding noise to their response, to allow for possible shifts in the the location of the preferred stimulus in morph space. As before, this procedure was repeated 100 times and the results were averaged to obtain the correct number of values.

Fig. 19 shows the population tuning properties of the resampled model distribution, for neurons with a noise level of $n = 0.08$ (a noise level of $n = 0.1$, as in Fig. 17, produced only slightly worse fits to the experimental distribution). We again find very good agreement with the experimental distribution, BWI and ROC distributions are not statistically significantly different ($p \geq 0.1$, Wilcoxon rank sum test), and the CCI distribution is only marginally different ($p = 0.03$).

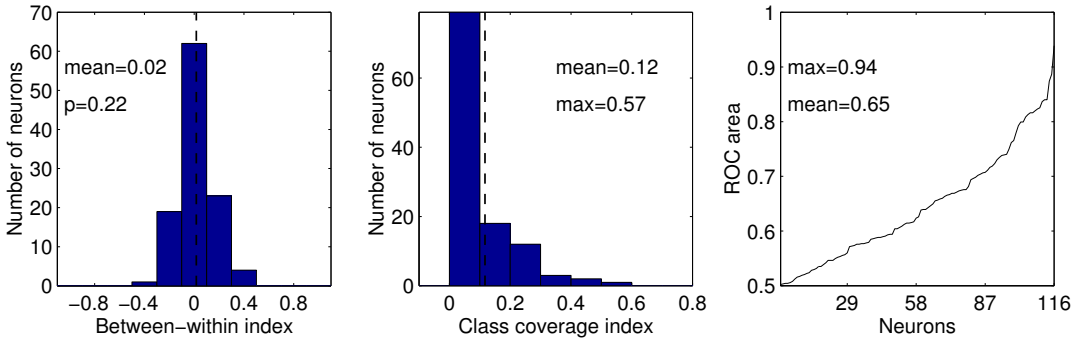

Figure 13: Experimental IT data. The plots show the distribution of BWI (left), CCI (center) and ROC (right) area values.

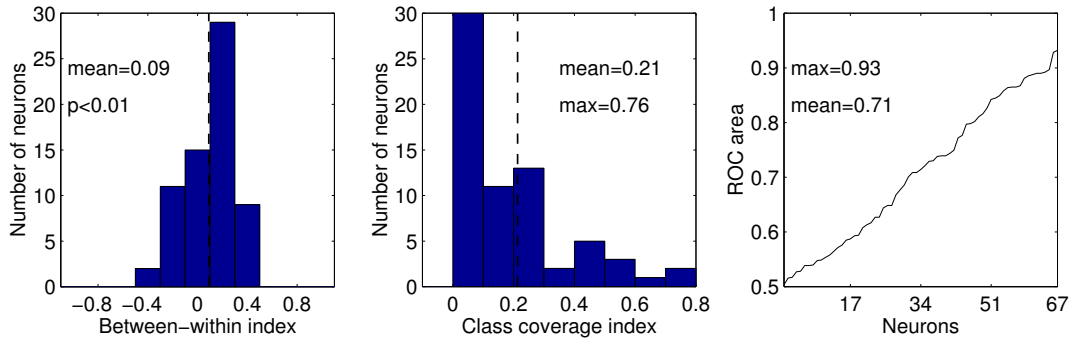

Figure 14: Experimental PFC data (sample period). The plots show the distribution of BWI, CCI and ROC area.

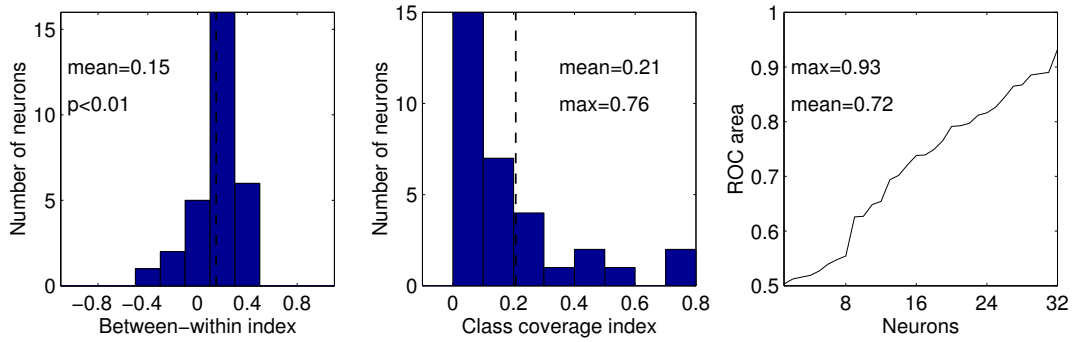

Figure 15: Experimental PFC data (delay period). The plots show the distribution of BWI, CCI and ROC area.

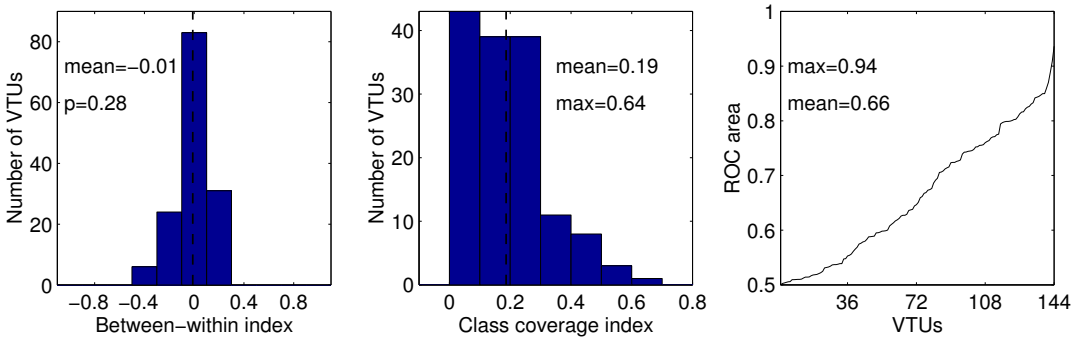

Figure 16: Model IT data for $a = 32, \sigma = 0.2$. The plots show the distribution of BWI, CCI, and ROC area values.

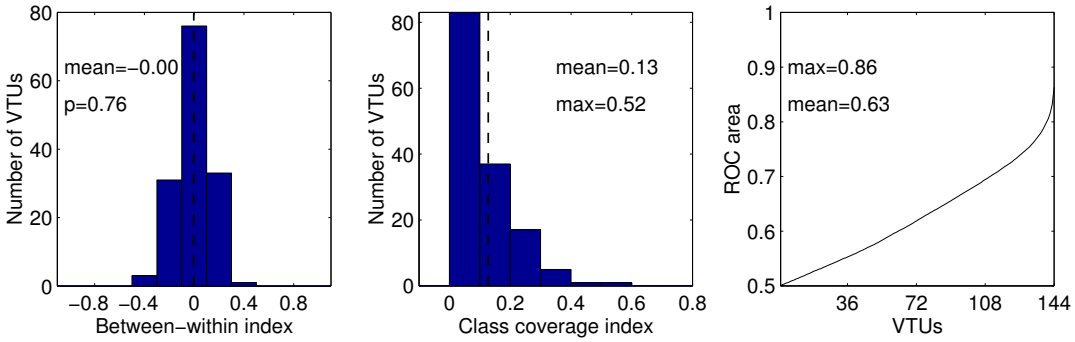

Figure 17: Model IT data for model units from Fig. 16, with added independent Gaussian noise of amplitude $n = 0.1$. The plots show the distribution of BWI, CCI, and ROC. Values shown are the average over 100 trials.

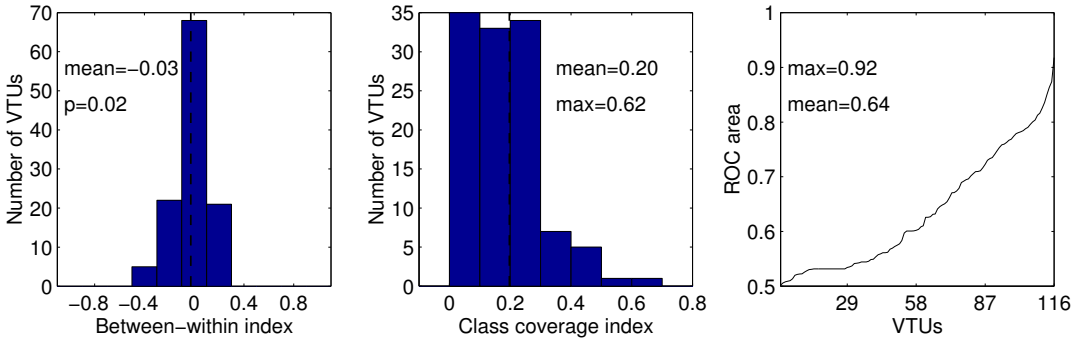

Figure 18: Resampled model IT data, deterministic units. The plots show the distribution of BWI, CCI and ROC area. Values shown are the average over 100 trials.

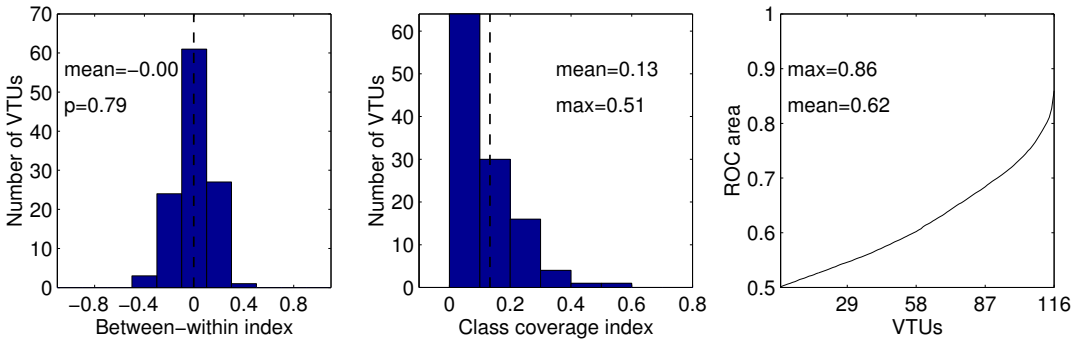

Figure 19: Resampled model IT data, noise level 0.08. The plots show the distribution of BWI, CCI and ROC area. Values shown are the average over 100 trials.

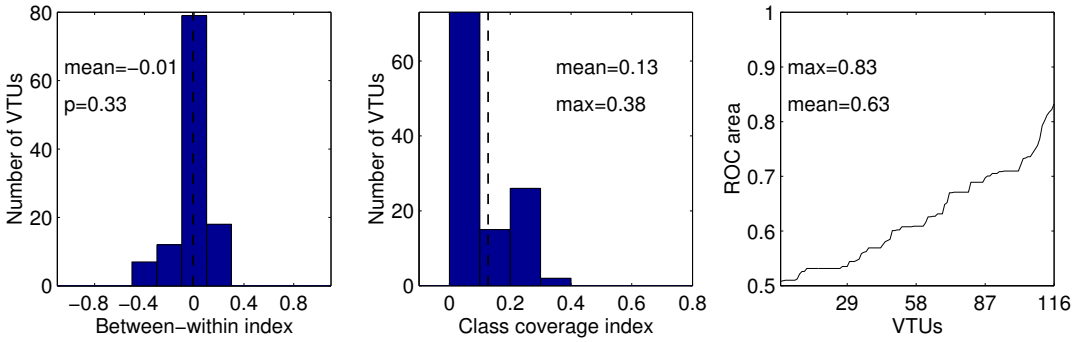

Figure 20: Fitted model IT data. The plots show the distribution of BWI, CCI and ROC area.

4.1.3 Fitting

As a further demonstration that model IT unit tuning is well compatible with experimental tuning, we have investigated how well the experimental IT population can be fitted using the model view-tuned units. To this end, we obtained the fitted population by selecting, for every cell found in the experiment, the best fitting model unit. This was done simply by adding and comparing the absolute difference of the BWI, CCI, ROC area, and morph index values for the experimental neuron and each model unit, respectively. Using this procedure produces a model unit population with tuning properties that are statistically not different from the population tuning found for the experimental IT neurons ($p > 0.1$ for all three of BWI, CCI, and ROC).

Thus, in summary, the degree of category tuning of experimental IT neurons appears to be very well captured by the population of view-tuned model units. As model units were trained without any explicit category information, the agreement of experimental IT and model data suggest that the learning of IT neuron response properties can be understood as largely driven by shape similarities in input space, without any influence of explicit category information.

4.2 Comparison of model units *vs.* PFC neurons

The PFC neurons show a BWI distribution with a positive mean significantly different from zero (sample period: 0.09, delay: 0.15), combined with higher average CCI values (sample: 0.21, delay: 0.21), with single neurons reaching values as high as 0.76 (sample and delay). Unlike in the IT case, this maximum value lies outside the range of CCI values of model units. Moreover, a positive average BWI of the magnitude found in the PFC data could only be obtained in the model with a significant number of of border-tuned neurons (cf. Fig. 5). Such border-tuned units have very low CCI values (cf. Fig. 7). CCI values of PFC neurons are higher than those of IT neurons, however. Thus, the tuning properties of PFC neurons *cannot* be explained in the model by mere stimulus tuning alone, but seem to require the influence of explicit category information during training.

5 Discussion

In this paper, we have analyzed the tuning properties of model view-tuned units tuned to the same stimuli that were used in a recent experiment [2, 3] in which monkeys were trained on a "cat/dog" categorization task, followed by recordings from their inferotemporal and prefrontal cortices. Using the same analysis methods as in the experiment, we found that view-tuned model units showed tuning properties very similar to those of monkey IT neurons. In particular, as with IT cells in the experiment, we found that some view-tuned units showed "categorization-like" behavior, *i. e.*, very high ROC values. Most notably, this tuning emerged as a consequence of the shape-tuning of the view-tuned units, with no influence of category information during training. In contrast, the population of PFC neurons showed tuning properties that could not be explained by mere stimulus tuning. Rather, the simulations suggest the explicit influence of category information in the development of PFC neuron tuning.

These different response properties of neurons in the two brain areas, with IT neurons coding for stimulus shape and PFC neurons showing more task-related tuning, are compatible with a recent model of object recognition in cortex [8, 10] in which a general object representation based on view- and object-tuned cells provides a basis for neurons tuned to specific object recognition tasks, such as categorization. This theory is also supported by data from another experiment in which different monkeys were trained on an identification and a categorization task, respectively, using the same stimuli [1], and which found no differences in the stimulus representation by inferotemporal neurons of the monkeys trained on different tasks. On the other hand, a recent experiment [12] reported IT neuron tuning emphasizing category-relevant features over non-relevant features (but no explicit representation of the class boundary, unlike in [3]) in monkeys trained to perform a categorization task. Further studies comparing IT neuron tuning before and after training on a categorization task or even different tasks involving the same set of stimuli, and studies that investigate the possibility of top-down modulation from higher areas (*e. g.,* PFC) during task execution, will be needed to more fully understand the role of top-down task-specific information in shaping IT neuron tuning. The present study demonstrated the use of computational models to motivate and guide the analysis of experimental data. Clearly, the road ahead will equally require a very close interaction of experiments and computational work.

Acknowledgements

We are grateful to Tommy Poggio and Earl Miller for useful comments and suggestions. Additional thanks to Marco Kuhlmann for technical assistance.

References

[1] de Beeck, H., Wagemans, J., and Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parametrized shapes. **4**, 1244–1252.

[2] Freedman, D. (2001). *Categorical Representation of Visual Stimuli in the Primate Prefrontal and Inferior Temporal Cortices*. PhD thesis, MIT, Cambridge, MA.

[3] Freedman, D., Riesenhuber, M., Poggio, T., and Miller, E. (2001). Categorical representation of vi-

sual stimuli in the primate prefrontal cortex. *Science* **291**, 312–316.

[4] Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cyb.* **36**, 193–202.

[5] Logothetis, N., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* **5**, 552–563.

[6] Macmillan, N. and Creelman, C. (1991). *Detection Theory: A User's Guide*.

[7] Riesenhuber, M. and Poggio, T. (1999). Are cortical models really bound by the "Binding Problem"? *Neuron* **24**, 87–93.

[8] Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* **2**, 1019–1025.

[9] Riesenhuber, M. and Poggio, T. (1999). A note on object class representation and categorical perception. AI Memo 1679, CBCL Paper 183, MIT AI Lab and CBCL, Cambridge, MA.

[10] Riesenhuber, M. and Poggio, T. (2000). Models of object recognition. *Nat. Neurosci. Supp.* **3**, 1199–1204.

[11] Shelton, C. (1996). *Three-Dimensional Correspondence*. Master's thesis, MIT, Cambridge, MA.

[12] Sigala, N. and Logothetis, N. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. **415**, 318–320.