



Detection and Identification of Rare Audiovisual Cues

Inesperata accident magis saepe quam quae speres.
(Things you do not expect happen more often than
things you do expect) Plautus (ca 200(B.C.))



Project no: 027787

DIRAC

Detection and Identification of Rare Audio-visual Cues

Integrated Project
IST - Priority 2

DELIVERABLE NO: D4.13
Modeling cortical cells with distance functions

Date of deliverable: 30.6.2009
Actual submission date: 27.7.2009

Start date of project: 01.01.2006

Duration: 60 months

Organization name of lead contractor for this deliverable: HUJI

Revision [1]

Project co-funded by the European Commission within the Sixth Framework Program (2002-2006)		
Dissemination Level		
PU	Public	
PP	Restricted to other program participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	X
CO	Confidential, only for members of the consortium (including the Commission Services)	



D4.13 - MODELING CORTICAL CELLS WITH DISTANCE FUNCTIONS

THE HEBREW UNIVERSITY OF JERUSALEM (HUJI)

Abstract:

Traditionally, the modeling of sensory neurons has focused on the characterization and/or the learning of input-output relations. Motivated by the view that different neurons impose different partitions of the stimulus space, we propose instead to learn the structure of the stimulus space, as imposed by the cell, by learning a cell specific *distance function*. Metaphorically speaking, this direction attempts to bypass the question of "how the cell speaks" (i.e. syntax), by focusing instead on the question of "what the cell says" (i.e. semantics).

Here we consider neural data from both the inferotemporal cortex (ITC) and the prefrontal cortex (PFC) of macaque monkeys in two distinct experimental settings. We learn a distance function over the stimulus space as induced by the cell response; the goal is to learn a function such that the distance between stimuli is large when the responses they evoke are very different, and small when the responses they evoke are similar.

Our main result shows that after training, when given new stimuli, our ability to predict their similarity to previously seen stimuli is significantly improved. As a direct consequence of this improvement we are able to predict the response of the cell to a novel stimuli using KNN over the learnt distances. The precision of this prediction as compared to the naive KNN prediction is shown to be improving as number of training samples increases. Furthermore, using our learned Kernel we obtain a partitioning of the stimuli space which is much more similar to the partition induced by the cell responses, and thus, are able in some of the cases to peek at the semantic partition induced by the cell.



Detection and Identification of Rare Audiovisual Cues

Inesperata accident magis saepe quam quae speres.
(Things you do not expect happen more often than
things you do expect) Plautus (ca 200(B.C.))



Table of Content

1	Introduction	1
2	Related Work	2
3	Experimental setting.....	3
	3.1 Problem formulation	3
	3.2 Neural computational setting implementation	4
	3.2.1 Representing the physical world	4
	3.3 Data representation.....	4
	3.4 Obtaining equivalence constraints from neural data	4
	3.5 Training a distance function.....	5
	3.5.1 Evaluation	5
4	Results: characterizing complex cells in the Visual System.....	5
	4.1 Results and evaluation	5
	4.1.1 Fitting Power and Generalization.....	5
	4.1.2 Generating response predictions	6
	4.1.3 Peeking into the semantics: Inferring iso-response manifolds.	8
	4.2 The story of a single cell.....	8
5	Conclusions and future work.....	9

1 Introduction

Functional characterization of single cells

Transformations of representations are an immanent part of our cognitive processes. Objects and sensations (with distinct physical properties) are analyzed, synthesized and transformed through the brain’s different pathways, taking detailed forms of membrane potentials and action potentials that eventually compose a larger mental representation¹. The question of what exactly is being represented in different brain regions has attracted great interest and some remarkable work, but while the picture is fairly clear in some domains, most remain in the dark.

A case of special interest is the characterization of sensory neurons: one of its main goals is to characterize dimensions in stimulus space to which the neurons are highly sensitive (causing large gradients in the neural responses), or alternatively dimensions in stimulus space to which the neuronal response is invariant (defining iso-response manifolds). This challenge is especially pronounced when trying to learn the representation of visual objects in higher brain areas, where simple feature representations (and models) are neglected in favor of complex, non-trivial and possibly semantic ones. Visual processing in the cortex is classically considered to be hierarchical, with simple feature representations being gradually neglected in favor of distributed complex object representations at the level of the inferotemporal cortex (ITC) [11, 15]. Consistent with this idea, a recent study [9] showed that it was possible to reliably readout position and scale invariant object category information from a small population of neurons (~ 200) in IT cortex. A second study pointed out a possible functional shift that might be taking place between the closely connected ITC and prefrontal cortex (PFC) regions [5, 6]. Others [2] point out the manifestation of non-linear representation already in the V4 regions and offers a model that may account for it. Understanding the actual manifestation of this shift and characterizing the role of single cells in such schemes as well as gathering information to support such models of representation has proved to be a major challenge.

The dominant approach to the functional characterization of sensory neurons attempts to learn the input-output relations $f : S \rightarrow R$ of a cortical neuron, and to predict the response of the neuron to novel stimuli. Prediction is often estimated from a set of known responses of a given neuron to a set of stimuli, by modeling some linear filter over the stimuli. One typical method of building such predictors uses linear models and their second order variants, in order to approximate the function that is assumed to generate the response [16, 12]. However, these models may fail when responses are highly non-linear or when the smoothness of the response as dependent on the stimuli space is lost [4] - a process which is hypothesized to occur as one moves along the Visual pathways from V4 to the ITC and further to the PFC [6].

The approach taken in this work avoids learning the input-output relation f . Instead, we attempt to learn the specific “geometric” structure induced by a neuron on the visual stimuli. In other words, we try to learn the non-linear wrapping of stimulus space as suggested by the neuronal responses.

Learning a single cell’s invariant space

Motivated by the view that different neurons impose different partitions of stimulus space which are not necessarily simply related to the simple feature structure of the stimuli [1], we attempt instead to learn the structure of the stimulus space by learning a *distance function*. Specifically, we characterize a neuron by learning a pairwise distance function over the stimulus space that is consistent with the similarities between the responses to different stimuli. A distance function is a function defined over pairs of data-points $D : S \times S \rightarrow \mathbb{R}$, which assigns a real (and possibly bounded) valued number

¹The dual representation terminology is in itself deeply embedded in the brain research discourse, see for example [18, 10]

to any pair of points from the input space $\{s_i, s_j\} \in S$. The assigned number measures the distance between pairs of points, which reflects the similarities between them.

Intuitively, a good distance function would capture the desired structure and assign small distance values to pairs of stimuli that elicit a similar neuronal response, and large values to pairs of stimuli that elicit different neuronal responses. Knowledge of the structure is in itself valuable, as it can be used to understand the kind of classification that a single cell performs on the stimuli space. Interestingly, it could also be used for prediction, using some variant of K Nearest Neighbors (KNN) with the learnt distance function.

Our approach offers several advantages: first, it allows us to aggregate information from a number of neurons and reach a good hypothesis even when the number of known stimuli responses per neuron is small, which is a typical concern in the domain of neuronal characterization. Second, unlike most functional characterizations that are limited to linear or weakly non-linear models, distance learning can approximate functions that are highly non-linear.

Metaphorically speaking, learning a cell specific distance function allows the investigator to bypass the question of "how the cell speaks", or "how many spikes are fired to a given stimuli?". Instead, we attempt to touch upon the more fundamental question of what exactly the cell is saying, or "what partition does the cell induce on stimulus space?".²

2 Related Work

Some neurophysiological background

The common view of sensory systems considers neurons as feature detectors arranged in an anatomical and functional hierarchy. In this view of the visual system, information from simple feature detectors along the retina converges at the level of the primary visual cortex (V1) to represent elaborate features of direction and orientation. This "strict hierarchy" view has recently been partially challenged, where the importance of context and background representation has been pointed out [13, 14]. However, the traditional approach is still to a large degree well accepted. Models of object recognition in the cortex have been mostly applied to tasks involving the recognition of isolated objects presented on blank backgrounds. Ultimately models of the visual system have to prove themselves in real world object recognition tasks, such as face detection in cluttered scenes, a standard computer vision benchmark task.

Serre et al. [13] have recently proposed an extension of an existing model of object recognition in the cortex, which characterizes the ventral visual pathway - from primary visual cortex (V1) to InferoTemporal cortex (ITC), a brain area thought to be crucial for object recognition. The model consists of a hierarchy of layers with two different types of pooling mechanisms: linear operations S , which build more complex features from simple ones, and nonlinear MAX pooling operations C , which increase the invariance of units to stimulus scaling and translation. The model explains how the so called view-tuned neurons in ITC can exhibit highly specific tuning to views of complex objects, while showing invariance to changes in stimulus position and scale.

These and other findings seem to suggest that highly complex representations of the environment cannot be accounted for by the mere use of simple linear models, and that a new approach to the functional characterization of these cells may be in order. A first step in this direction was taken by [17], introducing the learning of distance functions into this domain and showing some preliminary and promising results. Here we pursue this research direction further, trying to identify and characterize complex cells in the visual system. Our results may help to make this approach an accepted one in experimental neuroscience.

²A good distance function may be able to render the prediction task trivial, by using some kind of classification algorithm such as KNN.

3 Experimental setting

Distance function learning

While *distance function learning* is a somewhat new area of research, the concept of a distance function is well known and is widely used in various applications and for various computational tasks.³ Recent years have seen a lot of interest in distance function learning algorithms. These algorithms aim at automatically incorporating domain specific knowledge and/or side information into the applied distance function. In the general setting of such an algorithm, side information (typically in the form of equivalence constraints) is used to learn a pairwise function that adequately captures the structure of the space and the relations between the data-points.

Unlike the classical learning scenario, where one attempts to learn some function $f : X \rightarrow Y$ that approximates the relation between a given input and its resulting output using a training sample $S = \{(s_1, r_1), (s_2, r_2), \dots, (s_N, r_N)\}$, distance functions provide information about the similarity of pairs of points - essentially capturing relations within the input data-points themselves $D : S \times S \rightarrow \mathbb{R}$ rather than an input-output relation. In some cases capturing the relations between data-points can provide information which cannot be easily extracted from directly estimating input-output relations. Such cases occur when the input-output function is highly complex, multi staged, or just difficult to estimate. In other cases, the structure of the space is what one is interested in (rather than the transformation).

The case of single cell characterization is a case where both incentives for using distance function learning appeal: On the one hand estimating the actual transformation is hard and poses many physiological and technical limitations. On the other hand, we would argue, information about the partition of space may in many cases be much more interesting than predicting the actual response. Ideally, learning an adequate distance function may render the task of response prediction redundant.

3.1 Problem formulation

Our approach is based on the idea of learning a distance function over the stimuli space, using side-information extracted from the response space. The initial data consists of stimulus-response paired representations. To generate the side-information, we use an accepted similarity measure over pairs of points in the response space, which are used in turn to generate equivalence constraints on pairs of stimuli: two stimuli are related by a positive equivalence constraint if their paired responses are highly similar; they are related by a negative equivalence constraint if the paired responses are highly dissimilar.

Next, this side information is used to train an algorithm that learns a distance function between pairs of stimuli points, thus capturing implicitly the structure of the stimuli space as induced by the cell. Thus the cell becomes a teacher, specifying similarities between stimuli using its own language of action potentials. These similarities are then used to learn a cell-specific distance function over the space of all possible stimuli. This learned distance function should reveal what exactly is represented by the changes in the response of the specific cell.

We can formally define the computational task as follows:

Input: A set of stimuli-response pairs $\{s_i, r_i\}_{i=1}^N$

1. Represent the responses and stimuli in their own 'natural' feature space, along with an accepted similarity measure in the responses space.
2. Use the responses to extract equivalence constraints on stimuli, as described above.
3. Learn a distance function over the stimuli space $D(s_i, s_j) \rightarrow R$ using these constraints.
4. Use the generated distance function to understand and predict the nature of stimuli space.

³Much work has been done in the past which uses canonical distance functions, or alternatively uses hand-designed distance functions in various application domains. This is primarily due to the abundance of algorithms which are distance based - i.e., algorithms whose only input requirement is the pairwise distances between the input data-points, such as Nearest Neighbors (NN) methods and various clustering algorithms.

In the remainder of this section we present the details of our suggested scheme and how it is used for the characterization of visual neurons.

3.2 Neural computational setting implementation

The data was collected by Freedman et al [5, 6] and consisted of stimuli-response pairs of data recorded in the ITC and PFC of macaque monkeys. The stimuli used consisted in a continuous set of cat and dog stimuli constructed from six prototypes with three-dimensional morphing system. The stimuli were generated by morphing different amounts of the prototypes (see Fig. 1). This allowed to continuously vary the stimulus shape, and precisely define a category boundary. The category of a stimulus was defined by whichever category contributed more (50%) to a given morph. The behavioral paradigm required monkeys to release a lever if two stimuli (separated by a 1 sec delay) were from the same category (a category match), see [5, 6] for details. Overall 430 cells were identified and recorded in the ITC, out of which 162 (38%) were used in our analysis; 442 cells were identified in the PFC, out of which 61 (14%) were used in our analysis.

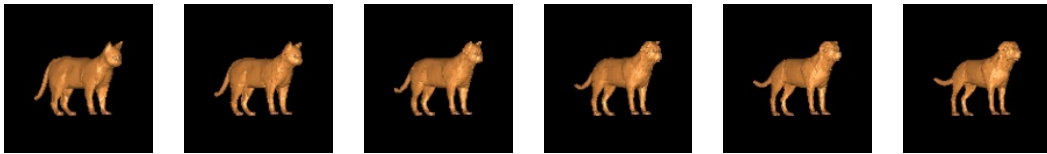


Figure 1: A morph line between 'Cat I' (left) and 'Dog I' (right). Three Prototypes of dogs and three prototypes of cats were morphed in six levels of inter-species 100%, 80%, 60%, 40%, 20%, 0% and 4 levels of within species morphing 100%, 60% 40% 0% yielding 42 Cat-Dog morphs and 12 within species morphs.

3.2.1 Representing the physical world

The question of how to represent the physical world is of crucial importance for any learning scheme. In fact, finding the "right" representation may render the whole problem trivial. In general our algorithm is designed to use any generic and well accepted representation, such that can be obtained via some simple manipulations of the image gray levels. In this paper we experimented with a physiological feasible representation as proposed by [?, 14], which eventually proved to be better for the task.

3.3 Data representation

As input for our learning algorithm we experimented with several representations, as hinted at above. We typically used the first 20 principal components of a representation in all scenarios, but the sparsity of training data in the leave-half- out scenario required further reduction to 15 principal components. The neuronal response for each stimulus was represented as a vector containing in each of its entries the spike rate of one out of multiple trials, where the number of trials per stimulus varied in the range 7-13.

3.4 Obtaining equivalence constraints from neural data

For the sake of simplicity, the distance between responses was measured using 2-way Anova between each vector of responses over multiple trials, thus creating a distance matrix over all pairs of responses. We then used the complete linkage algorithm to cluster the data into 8 clusters. All of the points in each cluster were marked as similar to one another, thus providing positive equivalence constraints. Negative constraints were determined to exist between points in the 4 furthest clusters.

3.5 Training a distance function

We took the scheme described above and implemented it using the Kernel-Boost distance learning algorithm described in [8],⁴ Kernel-Boost is a variant of the DistBoost algorithm [7] which was used in [17] and showed promising results. Kernel-Boost is a semi-supervised distance learning algorithm that learns distance functions using unlabeled datapoints and equivalence constraints. While the Dist-Boost algorithm has been shown to enhance clustering and retrieval performance, it was never used in the context of classification mainly due to the fact that the learnt distance function is not a kernel (and is not necessarily metric). Therefore it cannot be used by the large variety of kernel based classifiers that have shown to be highly successful in fully labeled classification scenarios. Kernel-Boost alleviates this problem by modifying the weak learner of DistBoost to produce a 'weak' kernel function. The 'weak' kernel has an intuitive probabilistic interpretation - the similarity between two points is dened by the probability that they both belong to the same Gaussian component within the GMM learned by the weak learner.

An additional important advantage of Kernel-Boost over Dist-Boost is that it is not restricted to model each class at each round using a single Gaussian model, therefore removing the assumption that classes are convex. This restriction is dealt with by using an adaptive label dissolve mechanism, which splits the labeled points from each class into several local subsets. An important inherited feature of KernelBoost is that it is semi-supervised, and can naturally accommodate unlabeled data in the learning process.

For each neuron, a subset of all pairs of stimuli was selected such that the responses of the two stimuli in a pair were either very similar or very dissimilar. The distance function was trained using a cross validation scheme to fit these constraints. The resulting distance functions generalized to predict the distances between the responses of a test stimulus and all trained stimuli.

3.5.1 Evaluation

We used a number of ways to evaluate the quality of the learned distance function. First, we evaluated the learned distance function by comparing the learned distances to the actual distances as measured over the cell responses,⁵ using rank correlation. These distances were used to cluster the stimuli data, which was then compared with the clustering induced by the cell responses using the rand index. Finally, this clustering was used along with a KNN classifier to generate predictions for novel samples in a cross validation scheme.

4 Results: characterizing complex cells in the Visual System

4.1 Results and evaluation

We narrowed our analysis to neurons which displayed some stimulus selectivity (Not necessarily category selectivity) . Such selectivity was established by performing n-way ANOVA with the 54 samples as a factor at $p < 0.01$. This analysis resulted in 162 ITC neurons and 61 PFC neurons. Hence forth we present results of evaluation on the ten most selective cell in each group.⁶ We start the results analysis by measuring the success of the learner to fit and generalize the distances as defines by the responses (Sec. 4.1.1) we then continue to show how this knowledge can be used for response prediction (Sec. 4.1.2) and stimuli classification (Sec. 4.1.3). We conclude this section by following closely the story if a single cell and by trying to infer it's semantics.

4.1.1 Fitting Power and Generalization

As a first step we examined the ability of our algorithm to fit the distances as induced by the cell responses. We evaluated this by measuring the mean rank correlation (Spearman) between the dis-

⁴In our comparative study Kernel-Boost performed extremely well, especially when given only a small amount of data.

⁵As stated before, how to compute the distance between two responses is a subject of debate; we used the widely used 2-way anova.

⁶Cells that did not display stimulus selectivity during the recording setting are a trivial case and are, thus omitted from the analysis. All reported results pertain to this sub selection of the data.

tances computed by our distance learning algorithm and those induced by the cell as measured by the 'real' distances over the responses. The main difference between the cell-induced distance matrix and the learnt one is that while the first can be calculated only on previously seen data the later can, once trained, be calculated on any two points. At this early stage of the experiment, a few different representations of the stimuli images were tested as input to our algorithm. The results as seen in 1 (top row) and Fig. 2(left most bar) show dramatic improvement after learning, regardless of the initial representation. This improvement persevered with any of six visual features combinations we tried (Not shown). We thus continued our analysis with the simplest representation at hand: a 20 PC representation over the gray scale representation of the images.

After establishing the improvement in fitting power across both data sets and all choices of representations, we turned to evaluate the generalization properties of the algorithm. We tested three scenarios: Leave-One-Out (LOO) where at each simulation one stimuli was left out of the training set for test only, Leave-Five-Out (LFO) and Leave-Ten-Out ((LTO) where a random sample of five and ten (in accordance) samples of the data was left out of the training set for later test (20 repetitions). For each stimulus that was tested in one of this manners, we measured its distances to all other stimuli using the learned distance function. We then computed the rank order Spearman correlation coefficient between the learned distances in the stimulus domain, and the 2-way anova distances between the appropriate responses. This procedure yielded a single correlation coefficient for each of the simulations using the stimuli which was left out. To measure performance, we took the average of the correlation coefficient over all runs for each cell.

	Distance Rank Correlation					
	ITC			PFC		
	Train	Test	Base	Train	Test	Base
Fitting	(0.63,0.01)		(0.19,0.01)	(0.57,0.01)		(0.18, 0.01)
LOO	(0.67, 0.01)	(0.43,0.01)	(0.19,0.01)	(0.60,0.01)	(0.33,0.02)	(0.14,0.01)
LFO	(0.66, 0.01)	(0.42,0.01)	(0.20,0.01)	(0.62,0.01)	(0.34, 0.02)	(0.13,0.01)
LTO	(0.67, 0.01)	(0.41,0.01)	(0.19,0.01)	(0.63,0.01)	(0.33,0.02)	(0.14,0.01)

Table 1: Distance correlation scores of different scenarios:bla baa

Results as seen in Fig. 2(upper row) show a persistent improvement in test correlation in all scenarios, as compared to baseline (starting) correlation. These results high light the face that the algorithm generalization properties are strong and some aspect of stimuli space topology was indeed captured. Test correlation scores are generally reduced as the size of sample used for training is reduced but the reduction is minor. (reduction of the error-bars is an artifact of the calculation method.) Cells also display very strong correlation between training performance and test performance. Fig. 2(bottom row) This results are achieved when averaging over all left out scenarios for a given cell, suggesting that a if a cell defines some partition on stimuli space then on average this partition is captured, however as can be deduced for the large error bars, this does not imply succeeding on each left out stimuli, we assume that good results depend on good sampling of stimuli space.

4.1.2 Generating response predictions

Having proven the generalization properties of the learning algorithm, we attempt to use the newly learnt distance function to generate response prediction to novel stimuli. We do that choosing for a novel stimuli the 1st -Nearest Neighbor (1-NN) in the training sample. and predicting that the cell would respond to the novel stimuli in a similar way to the response of that same cell to the closest (by our new definition of closeness) stimuli. Translating the learnt distance function into prediction is not an easy task: As the stimuli space is very restricted in this experimental setting, responses tend to be relatively smooth locally, in the sense that similar (gray level) stimuli tend to respond stimuli Fig. ?? . Our prediction scheme manages to out perform both linear regression prediction and naive 1-NN on the gray levels of the stimuli. The success in prediction, and our advantage over the alternative predictors is highly dependent on the size of the training sample and increases as the sample size increases, which is consistent with the increase in test distance correlations demonstrated earlier. We do better or equal to other predictors in 64.44%, 66.2%, 61.10% for ITC cells. and 68.03% 65.70% and 61.55% for PFC cells with 53, 49, 44 training samples in accordance.

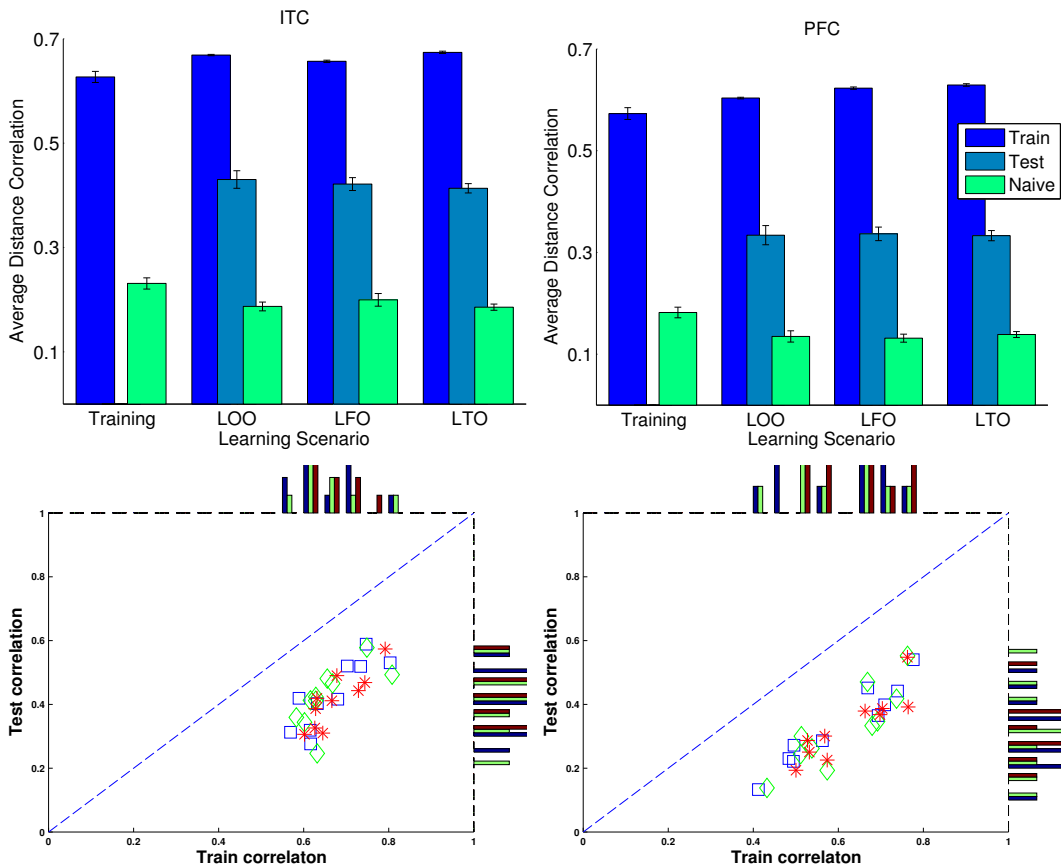


Figure 2: **Distance correlation score** 1) Trained vs. Test vs. Naive performance (upper row): The algorithm achieved a dramatic increase in average rank correlation, suggesting that some aspect of the structure of stimuli space was indeed captured. 2) Train vs. Test correlations (bottom row): correlation of train test was 0.83, 0.82, 0.95 on ITC cells. and 0.82, 0.94, 0.92 for PFC cells for LOO, LFO and LTO in accordance.

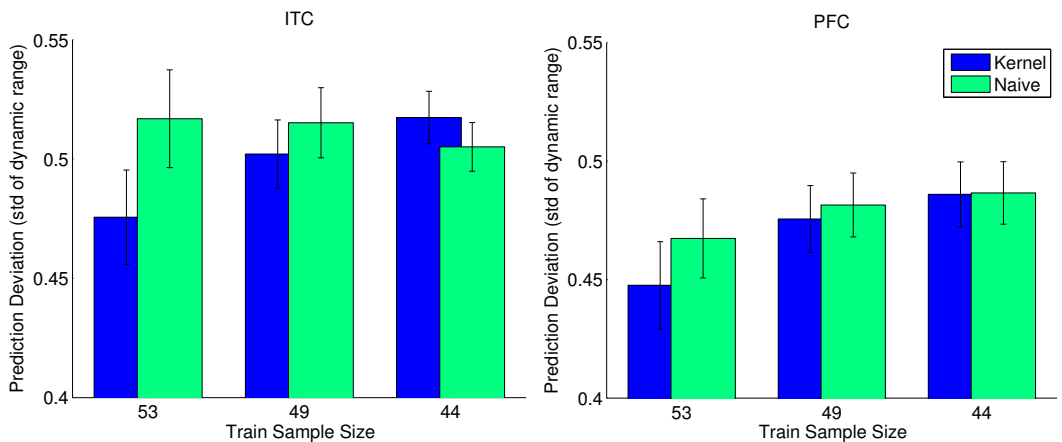


Figure 3: **Prediction advantage**: shown above are the percentage of simulations that achieved advantage in prediction power. An advantage in predictive power is calculated as smaller absolute distance to actual response. Our algorithm is advantageous in the majority of the scenarios, and the advantage becomes stronger as sample size increases.

4.1.3 Peeking into the semantics: Inferring iso-response manifolds

To a degree of abstraction a single cell divides the world it sees to a few iso-response manifolds. This division is reflected by a distinct response pattern to each of these manifolds. In our setting, for the sake of simplicity, we assume that a cell induces a binary partition on the stimuli space and thus defines two iso-response manifolds. Inferring iso-response manifolds is crucial for an experimental setting, and can lead a researcher through new interesting unexperimented pathways. As our algorithm is learning a kernel over stimuli space we can use this kernel in an SVM [?] classifier to better classify novel stimuli into one of the two manifolds. The fact that a kernel can help in better classifying stimulus space has been highlighted before. We show here that if we were to have the perfect Kernel - that is the response kernel. (Which we obviously don't have in an experimental setting) we would have been able to guarantee a 95% classification accuracy on unseen data. This further strengthens the notion that trying to approximate this kernel is a worthwhile assignment. To experiment with this notion we used the LIBSVM library implementation [3] where we trained an SVM with a response Kernel⁷, a linear Kernel, a quadratic Kernel, an RBF Kernel and our own learnt Kernel. We repeated the training and testing procedure using 44 (LTO), 49 (LFO), and 53 (LOO) samples and averaged the agreement of classification to the one defined by the cell.

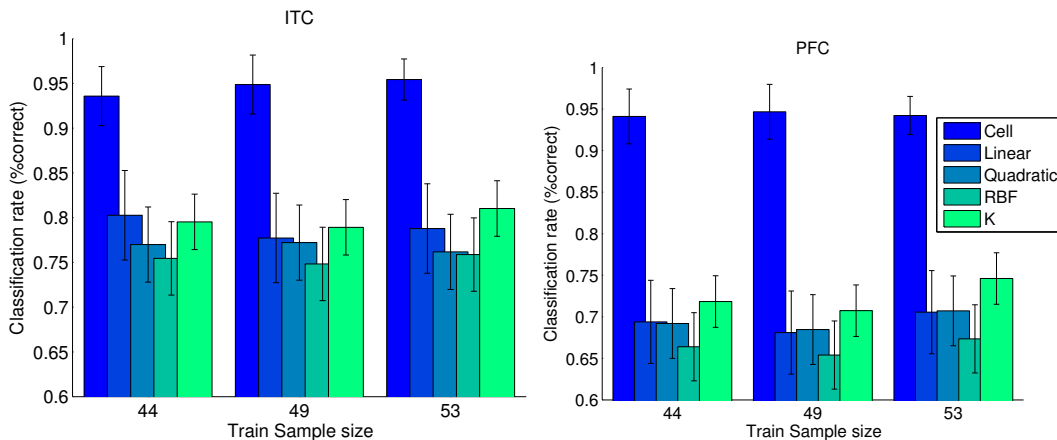


Figure 4: **Classification performance :performance of an SVM binary classifier in different learning scenarios:**The learnt Kernel (green bar) improves classification rate to response no response classes as induced by the cell as compared to classic Kernels (Linear, Quadratic and RBF). Note that having the "correct" Kernel - that is the response Kernel (left most bar) insures an almost perfect classification.

This advantage in SVM classification performance though small hides a much drastic improvement in the way the input space can now be understood. This is revealed through 2-D embedding of the new learnt distance matrix. As an example we show two examples of cells in a leave ten out setting. We took the distance matrix as learnt by the algorithm and embed in 2-D. This enables us to evaluate visually the effect of learning on our ability to identify partitions in stimuli space which are far from being evident in the gray scale relations.

4.2 The story of a single cell

We start the presentation of our results with an illustrative single cell example. In the scenario at hand we have gathered some information about the cell's responses to a sample of stimuli. To capture what exactly is being represented by a single cell, each stimulus of the training sample was clustered to N clusters according to the responses it elicited) for simplicity and as a good abstraction we assume the cell divides the world into two iso-response manifold: 'Response' / 'No Response'. Next, we use our algorithm learn a distance function for the cell using equivalence constraints extracted from the previously done clustering. We can then use the learnt distance matrix to re-cluster the unseen stimuli. Note that the cell does not partition the space into a simple gray-scale related features, but

⁷We employ the so-called empirical kernel map see [?] to convert the response difference score to a kernel.

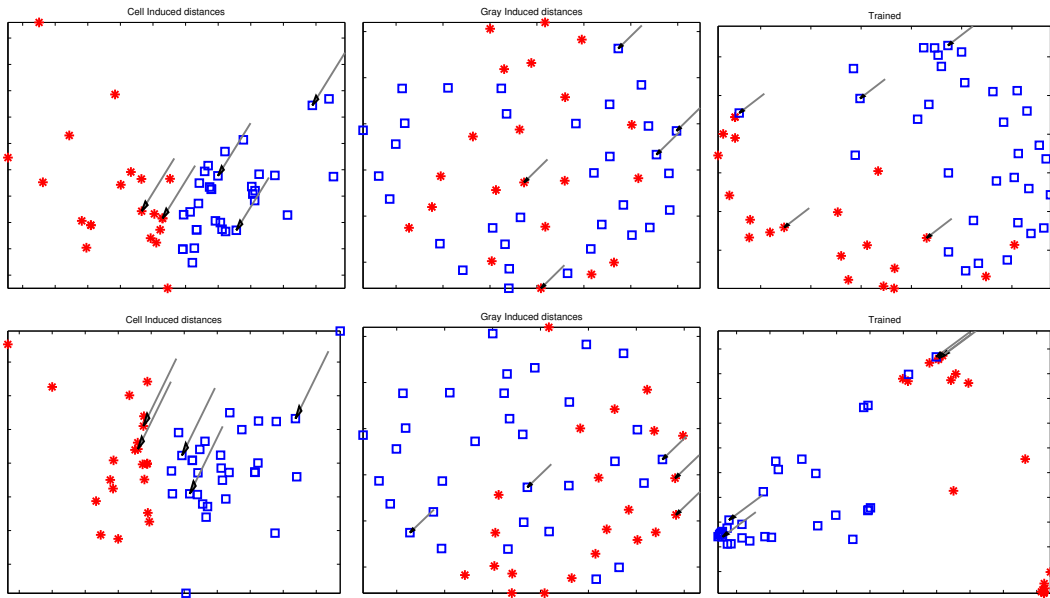


Figure 5: **Visualizing classification effect:** Two cells were selected to display the effect of learning on the embedding of points in 2-D. For each cell we show response distances based embedding (left column) gray scale based embedding (middle column), Trained based embedding (right column). In each of the cases we used MDS to embed the data points in 2-D. Points are partitioned according to the response they elicit to two iso-response manifolds. (red-circle vs. blue-square) The goal of the learner is to be able to replicate the partition induced by the cell. Numbered points were left out during the training phase

typically induces a more complicated partition. As shown in Fig. 6, this partition is captured almost perfectly by our algorithm (bottom matrix) panel).

5 Conclusions and future work

Our proposed scheme can serve as an integral part in neuroscientist's experimental setup. An effective distance function learning machinery should be able to direct the researcher towards "interesting" areas in the stimuli space as defined by the cell itself, to reduce the time and frustration involved in a search based on trial and error. For that to happen, we plan to develop the scheme so that it will work in an online manner, being able to handle information fast and in an accumulative manner.

References

- [1] O. Bar-Yosef, Y. Rotman, and I. Nelken. Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. *J Neurosci*, 22(19):8619–8632, Oct 2002.
- [2] C. Cadieu, M. Kouh, A. Pasupathy, C. E. Connor, M. Riesenhuber, and T. Poggio. A model of v4 shape selectivity and invariance. *J Neurophysiol*, 98(3):1733–1750, Sep 2007.
- [3] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [4] S.V. David and J. L. Gallant. Predicting neuronal responses during natural vision. *Network*, 16(2-3):239–260, 2005.
- [5] D. J. Freedman, M. Riesenhuber, T. Poggio, and E. K. Miller. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, 291(5502):312–316, Jan 2001.
- [6] D. J. Freedman, M. Riesenhuber, T. Poggio, and E. K. Miller. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J Neurosci*, 23(12):5235–5246, 2003.
- [7] T. Hertz, A. Bar-Hillel, and D. Weinshall. Learning distance functions for image retrieval. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 2004*, volume 2, pages II–570–II–577 Vol.2, 2004.

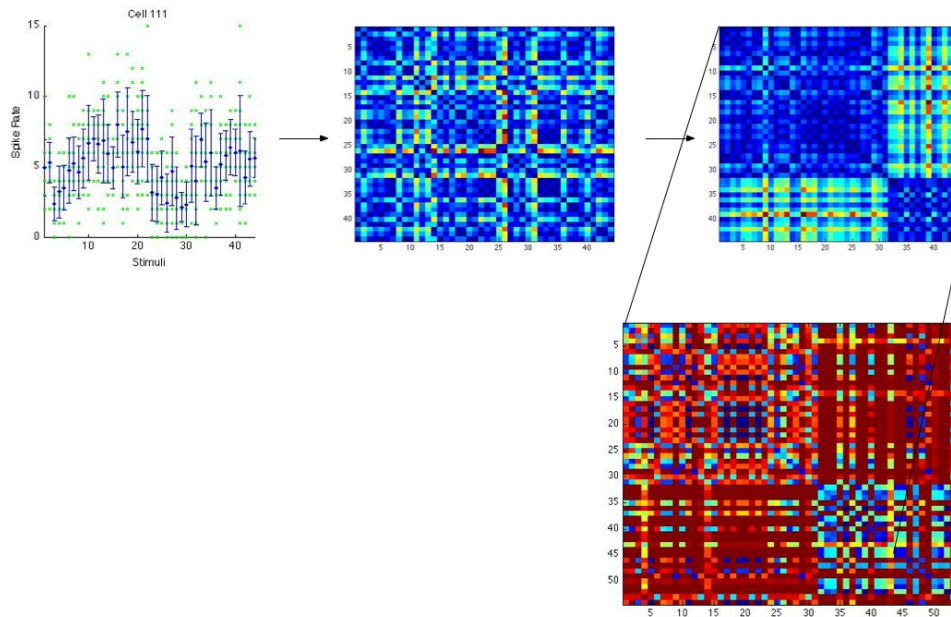


Figure 6: **Single cell example** A selection of the stimuli (44/54 was used as training sample (top left) to extract constraint based on clustering in the response space. (top middle and right) We then learned a distance function that represents a hypothesis about the partition induced by the cell on the stimuli space.

- [8] T. Hertz, A. Bar-Hillel, and D. Weinshall. Learning a kernel function for classification with small training samples. In *International Conference on Machine Learning (ICML)*, 2006.
- [9] C. Hung, G. Kreiman, T. Poggio, and J. DiCarlo. Fast read-out of object identity from macaque inferior temporal cortex. *Science*, 310:863–866, November 2005.
- [10] H. Imamizu, S. Miyauchi, T. Tamada, Y. Sasaki, R. Takino, B. Ptz, T. Yoshioka, and M. Kawato. Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature*, 403(6766):192–195, Jan 2000.
- [11] N. K. Logothetis and D. L. Sheinberg. Visual object recognition. *Ann. Rev. Neurosci.*, 19:577–621, 1996.
- [12] N. C. Rust, O. Schwartz, J. A. Movshon, and E. P. Simoncelli. Spatiotemporal elements of macaque v1 receptive fields. *Neuron*, 46(6):945–956, Jun 2005.
- [13] T. Serre, M. Kouh., C. Cadieu, U. Knoblich, G. Kreiman, and T. Poggio. A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex. AI Memo 2005-036 / CBCL Memo 259, MIT, Cambridge, MA, 2005.
- [14] T. Serre, A. Oliva, and T. Poggio. A feedforward theory of visual cortex accounts for human performance in rapid categorization. *Proceedings of the National Academy of Science*, 104(15):6424–6429, 2007.
- [15] K. Tanaka. Inferotemporal cortex and object vision. *Ann. Rev. Neurosci.*, 19:109–139, 1996.
- [16] F. E. Theunissen, K. Sen, and A. J. Doupe. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci*, 20(6):2315–2331, Mar 2000.
- [17] I. Weiner, T. Hertz, I. Nelken, and D. Weinshall. Analyzing auditory neurons by learning distance functions. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 1481–1488. MIT Press, Cambridge, MA, 2006.
- [18] D. M. Wolpert, Z. Ghahramani, and M. I. Jordan. An internal model for sensorimotor integration. *Science*, 269(5232):1880–1882, Sep 1995.