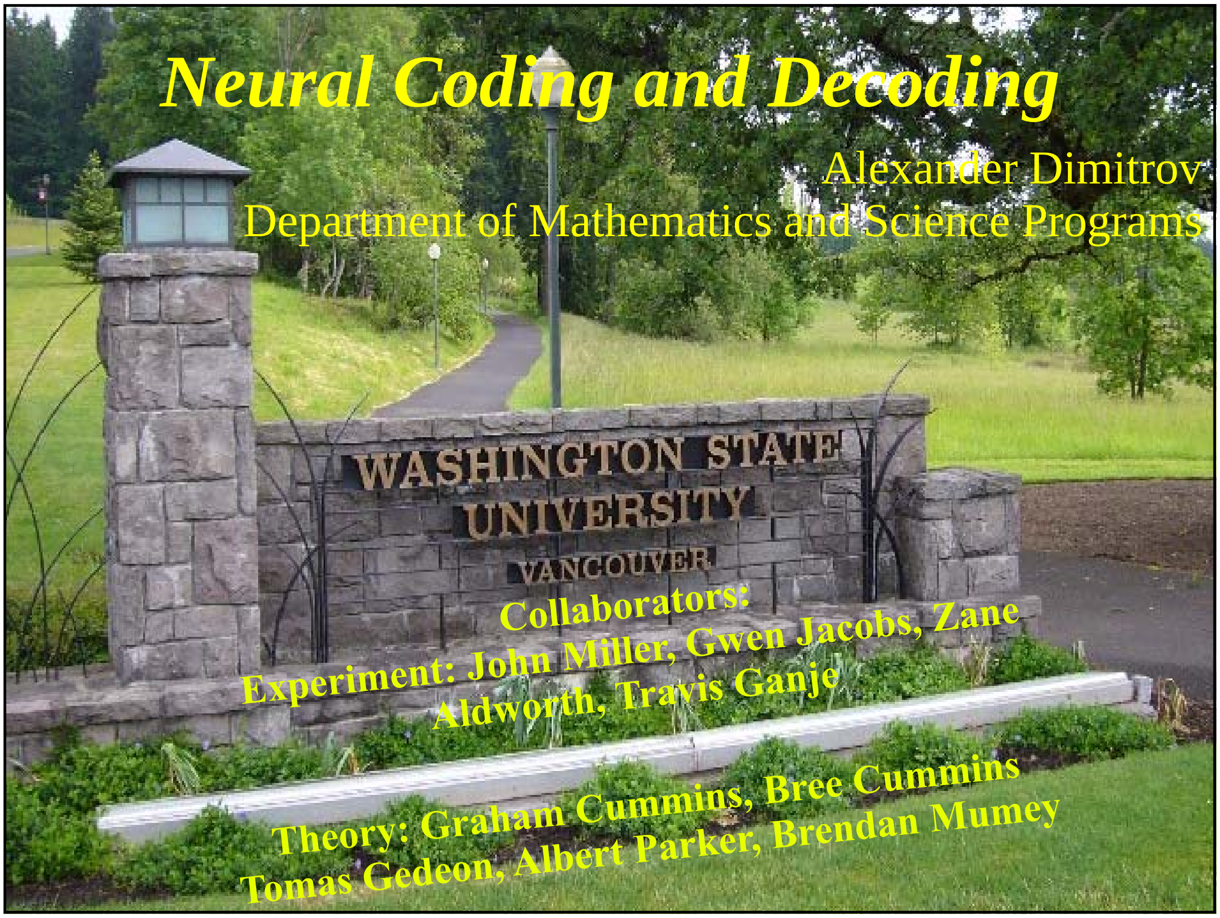
10-29-2010

Neural Coding and Decoding

Alexander Dimitrov

Washington State University

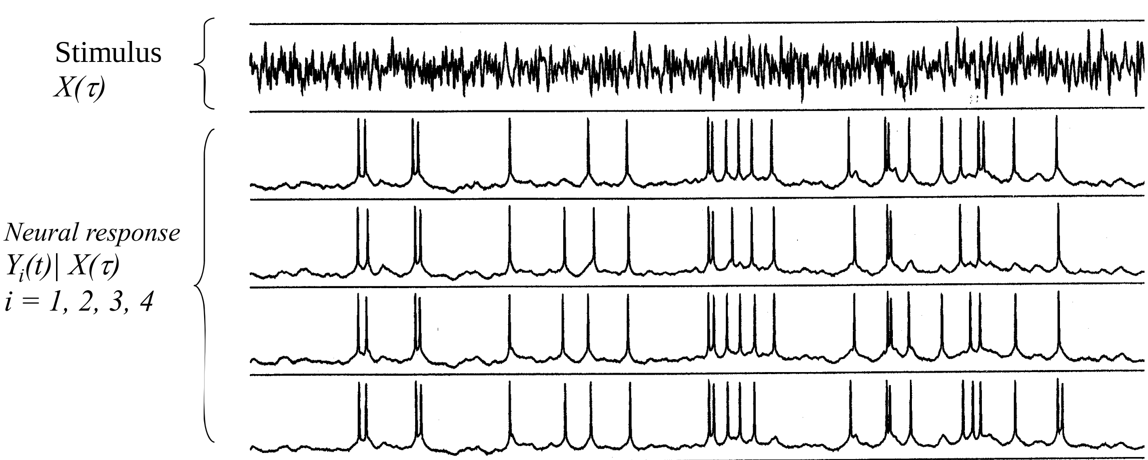


How does neural activity represent information about stimuli from the environment?

Summary of important features of neural codes

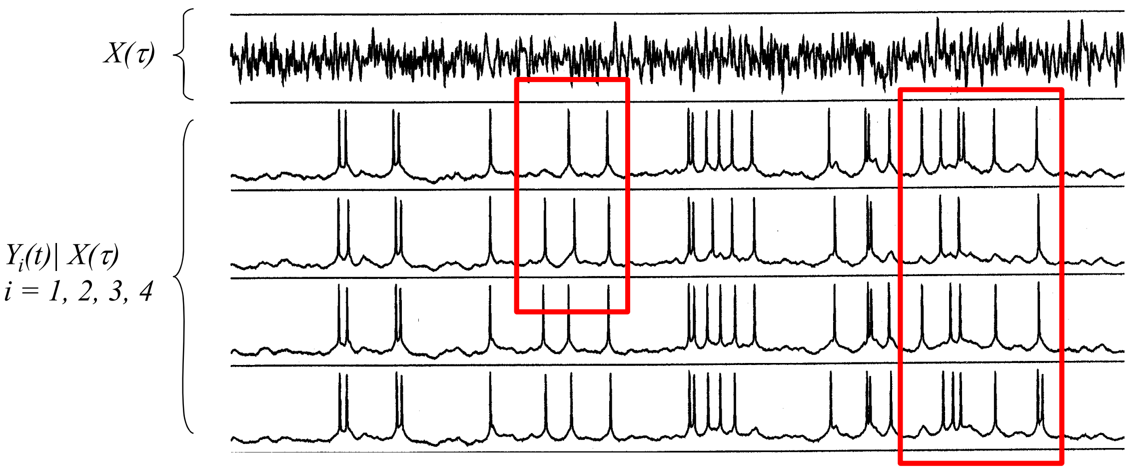
The "same" stimulus has to be recognized as such …

Given a stimulus (*e.g., sound waves*), we and the organism observe the response associated with it (*here, spike trains*):



… but the dictionary is not *deterministic*!

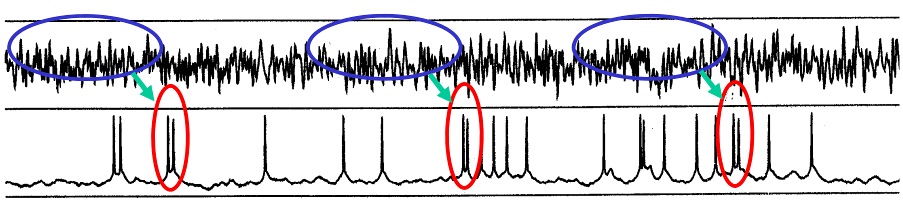
Given a stimulus, we observe many slightly different neural responses (*noisy spike trains*):



Neural coding is *stochastic*!!

… and the response is seemingly not consistent!

The "same" response (*sequence of spikes*), may be associated with many different sensory stimuli:



Neural Coding and Decoding

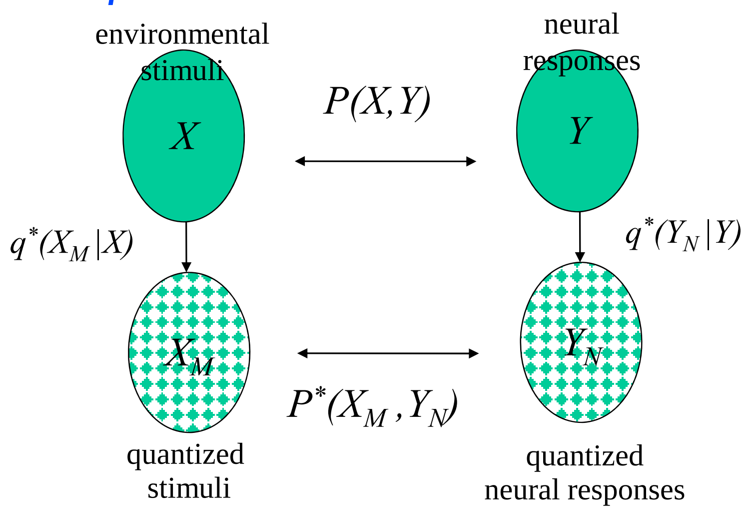
The Problem: Determine a *coding scheme:* How does neural activity represent information about environmental stimuli?

Summary of code feature we consider important:

- Any animal perceives its environment only by observing its own internal representation through neural activity.
- The code must deal with uncertainties introduced by the environment and neural architecture. Activity is by necessity *stochastic* at this finer scale.
- An animal needs to recognize the same object on repeated exposures. Failures at this stage may endanger its well-being. Coding has to be mostly *deterministic* at this level.
- Ecological reasons for uncertainty – what if the code **was** deterministic? Somebody else can break it!

Recovering a coding scheme

Goal: find P(X,Y).
Strategy: Determine the correspondence, P^*, between reproductions (X_M, Y_N) of few elements, such that P^* preserves as much "relevant" P as possible.
Discard details of P that "don't matter".



$$P(x_m, y_n) = \sum_{x,y} q(x_m|x)\,q(y_n|y)\,P(x,y)$$
$$= E_{P(X,Y)}\,q(x_m|x)\,q(y_n|y)$$

Optimal quantization

We define "relevant" as preserving the mutual information *I(X;Y)* in between stimulus and response.

New Goal: Find the quantizers q^* that minimize

$$D_I = I(X;Y) - I(X_M;Y_N)$$

for fixed M and N. The same as minimizing effective distortion $D_{eff} = -I(X_M, Y_N)$.

Information Theory

The Foundation of the Model

- The structure underlying information theory is a probability measure space (source, random variable). An *expectation* E_X is an integral over the probability measure.
- A signal *x* is produced by a source (r.v.) *X* with a probability *p(X=x)*. A signal *y* is produced by another source *Y* with probability *p(Y=y).*
- A ***communication channel*** is a relation between two r.v.'s *X* and *Y*. It is described by the conditional probability *Q(Y | X).*
- ***Entropy:*** the uncertainty, or self information of a r. v.

$$H(X) \equiv E_X\left(\log\frac{1}{p(X)}\right) \equiv \int\left(\log\frac{1}{p(X)}\right)p(X)dX$$

- ***Conditional Entropy:*** the reduced uncertainty of one r.v, if another is observed.

$$H(X|Y) \equiv E_{X,Y}\left(\log\frac{1}{Q(X|Y)}\right)$$

- ***Mutual Information:*** the amount of information that one r.v. contains about another

$$I(X;Y) \equiv E_{X,Y}\left(\log\frac{p(X,Y)}{p(X)p(Y)}\right)$$

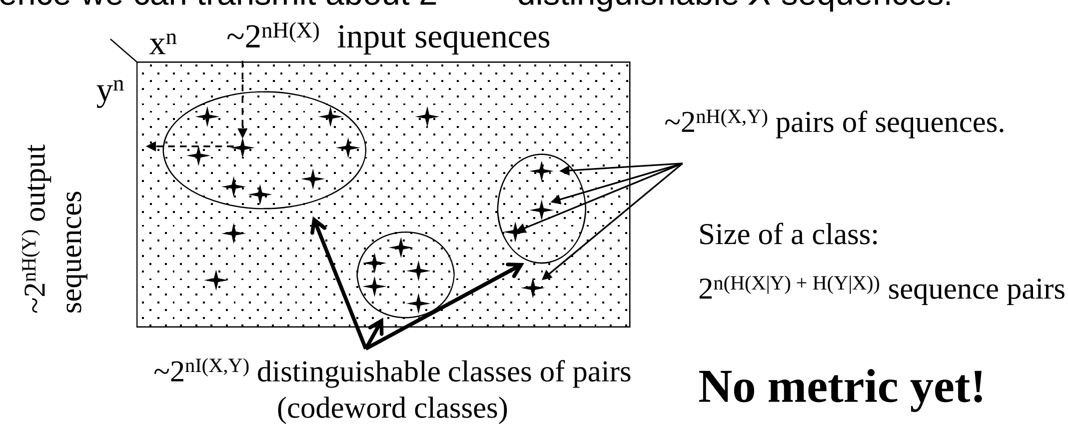
Transmission of information

Consider the encoding process in a probabilistic framework. Information theory makes a few important statements on how messages can be transmitted.

- A signal $x \in X$ is produced by a source with a probability *p(x)*. A source is characterized by its entropy *H(X):* it can be described completely using no more than *H(X)*+1 bits per symbol on the average.
- A channel is a relation between two random variables *X* and *Y*. It is completely described by the conditional probability q(y|x). A channel is characterized by its capacity $C = \max_{p(x)} I(X;Y)$.
- Sending data through noisy channels: the joint source channel coding theorem. *A finite alphabet process with entropy rate* H(X) *can be transmitted through a channel with capacity* C *with vanishingly small probability of error iff* H(X) < C.

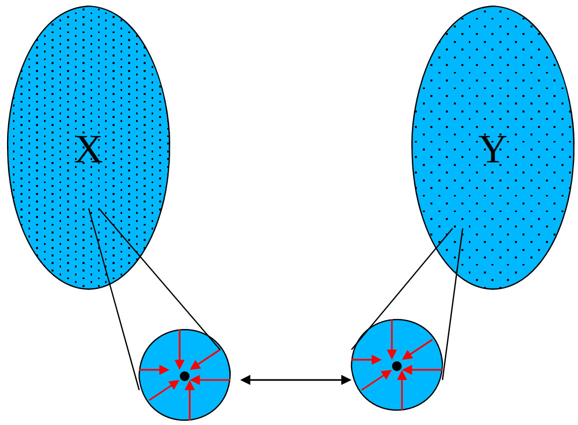
The structure of a communication channel.

- The total number of possible (high probability; typical) output sequences is about $2^{nH(Y)}$.
- For each input sequence x^n there are about $2^{nH(Y|X)}$ possible sequences in Y.
- So that no two *X*-s produce the same Y, the output should be divided in subsets of size about $2^{nH(Y|X)}$, corresponding to different input *X*-s.
- The total number of disjoint sets then is about $2^{n(H(Y) - H(Y|X))} = 2^{nI(X,Y)}$. Hence we can transmit about $2^{nI(X,Y)}$ distinguishable *X* sequences.



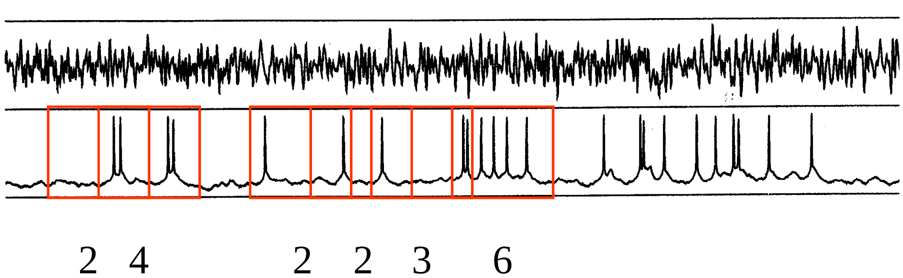
No metric yet!

Quantization



A quantization example

- Rate coding



Quantization

A source Y *can be related to another random variable* Y_N *through* the process of *quantization (lossy compression).* Y_N *is referred* to as the *reproduction of* Y. *The process is defined by a map*

$$q(Y_N|Y) : Y \to Y_N$$

called a *quantizer*. In general, quantizers can be stochastic*:* q assigns to $y \in Y$ *the probability* that the response y belongs to an abstract class $y_N \in Y_N$. *A deterministic quantizer* (simple function) is a special case in which q takes the values of 0 or 1 only. It can be shown that the mutual information $I(X; Y)$ is the least upper bound of $I(X_M; Y_N)$ *over all possible reproductions* $(X_M; Y_N)$ *of* $(X; Y)$. Hence, the original mutual information can be approximated with arbitrary precision using carefully chosen reproduction spaces.

Rate Distortion Theory

Rate distortion theory is concerned with reduced representations of random variables (lossy compression). The quality of reproduction (fidelity) is assessed through a *distortion function.*

Consider the quantization $X \to X_N$.

Definition 3. *A (pointwise) distortion function, or distortion measure is a mapping*

$$d : X \times X_N \to R^+$$

from the set of source/reproduction pairs into the set of nonnegative reals. The distortion is a measure of the "error" made by representing the symbol $x \in X$ *with* $x_n \in X_N$.

Example 3 (Squared error distortion). $d(x; x_n) = (x - x_n)^2$.

Definition 4 (Expected (mean) distortion function).

$$D(X;X_N) = E_{p(x;xn)}d(x;\, x_n)$$

Definition 5 (Rate distortion problem). *The information rate-distortion function R(D) for a source X with a distortion measure* $d(x;\, x_n)$ *is defined as*

$$R(D) = \min_{q(x_n|x):D(X;X_N)\le D} I(X;X_N)$$

where the minimization is over all conditional probabilities $q(x_n|x)$ *for which the joint distribution* $p(x;\, x_n) = q(x_n|x)p(x)$ *satisfies the expected distortion constraint. Equivalently, one may consider the distortion-rate problem*

$$D(R) = \min_{q(x_n|x):I(X;X_N)\le R} D(X;X_N)$$

Quantization theory.

The quantized information quantities in Y_N are (*Gray '94)*

$$I(X_M;Y_N) \equiv \sum_{i,j}^{M,N} p(x_i, y_j)\ln\frac{p(x_i, y_j)}{p(x_i)p(y_j)} \quad H(X_M) \equiv \sum_{i=1}^{M} p(x_i)\ln\frac{1}{p(x_i)}$$

If a quantizer h refines f then

$$\begin{array}{ccc} Y \xrightarrow{h} Y_M \xrightarrow{f} Y_N & & \text{(Markov relation)} \\ H(Y_M) & \ge & H(Y_N) \\ H(X_L \mid Y_M) & \le & H(X_L \mid Y_N) \\ I(X_L;Y_M) & \ge & I(X_L;Y_N) \end{array}$$

$I(X\,;Y) = \sup I(X_M\,;Y_N)$.

The way we build up measures in measure spaces.

Approaches

Constrained optimization: Search for the quantizers (conditional probabilities, $\Sigma q=1$) $\{q^*(X_M|X)\ ,\ q^*(Y_N|Y)\ \}$ that maximize the entropy,

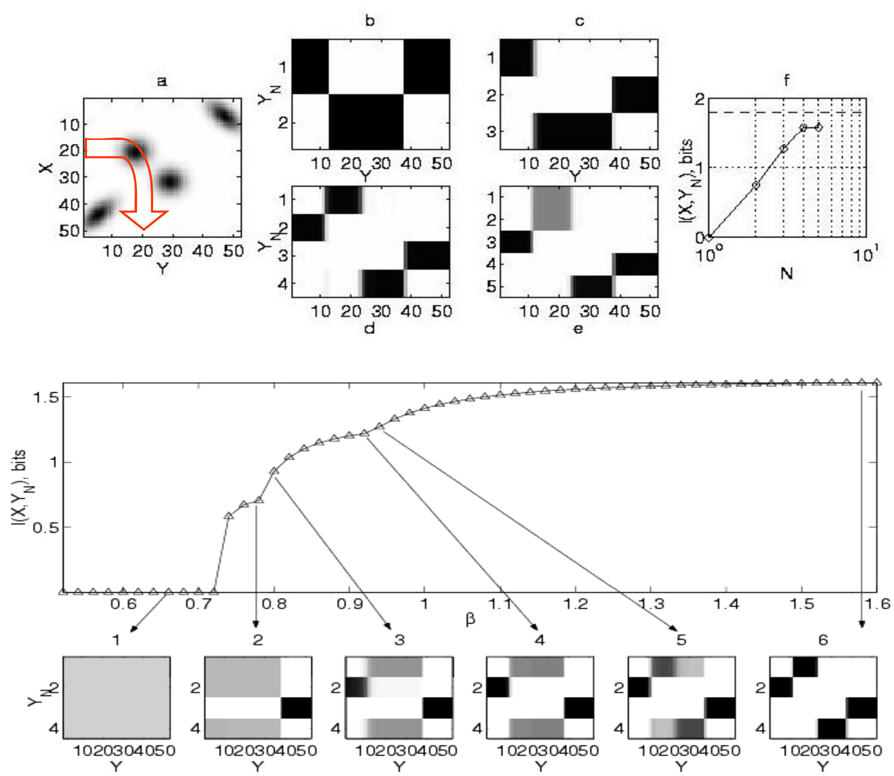
$max\ H(X_M, Y_N\,|X,Y)$ constrained by $I(X_M, Y_N\,) \leq I_o$ and let $I_o \rightarrow I_{max}$

Annealing: I_o above is a parameter anyway, so maximize the parametric cost function

$max\ H + \beta\, I$, vary β.

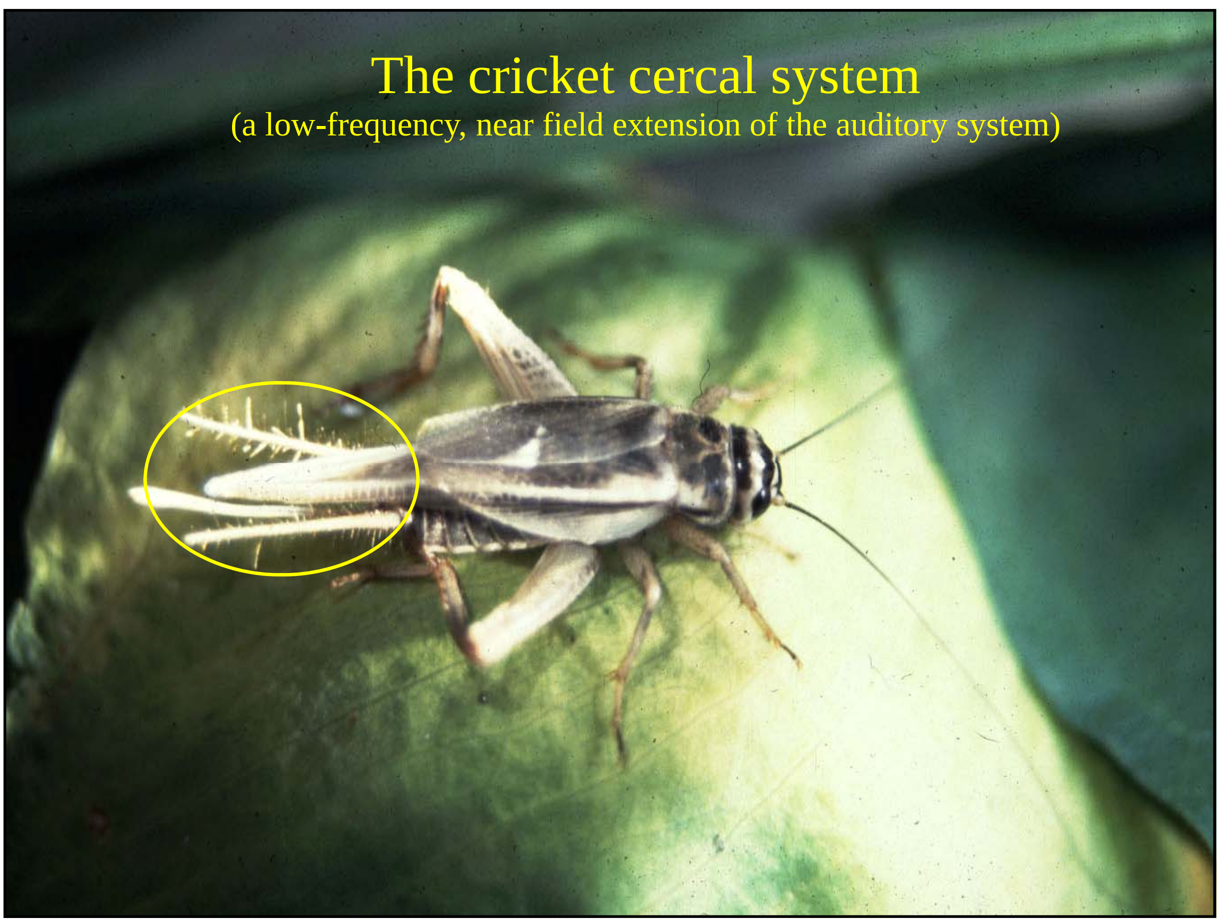
$\beta = 0$, purely max H; $\beta \rightarrow \infty$, max I.

Simulations

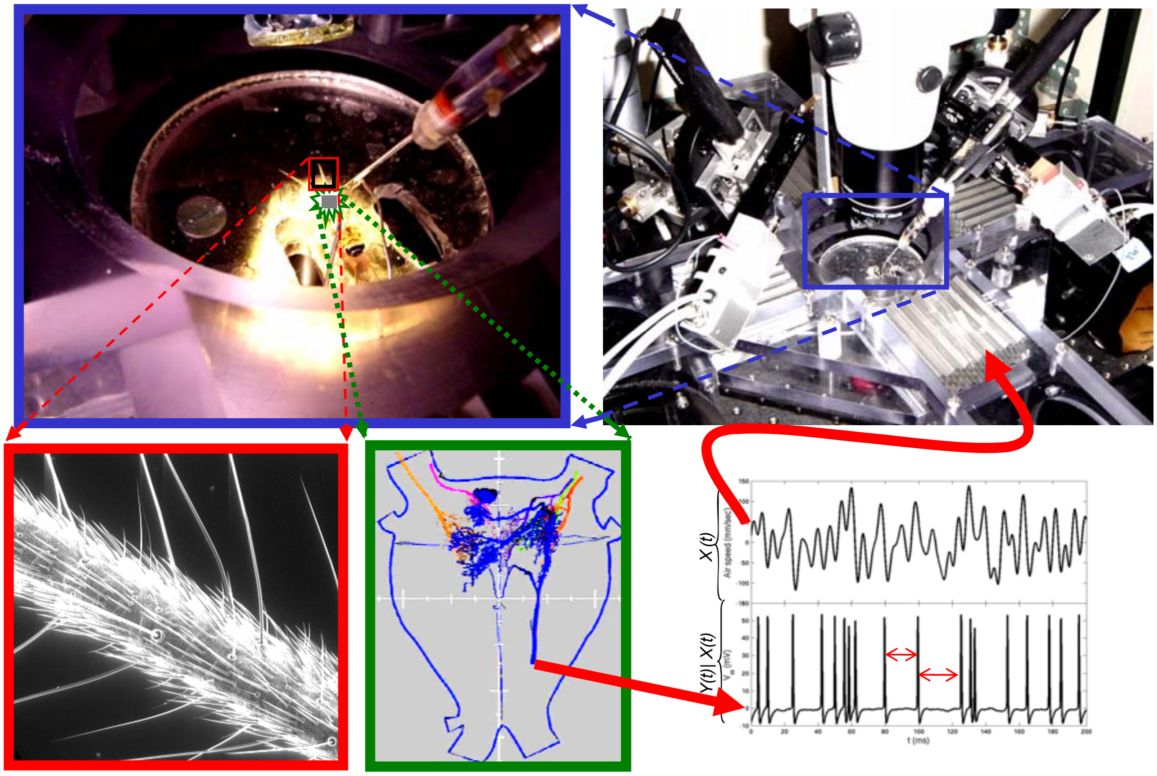


The sensory system challenge: dealing with complex stimuli.

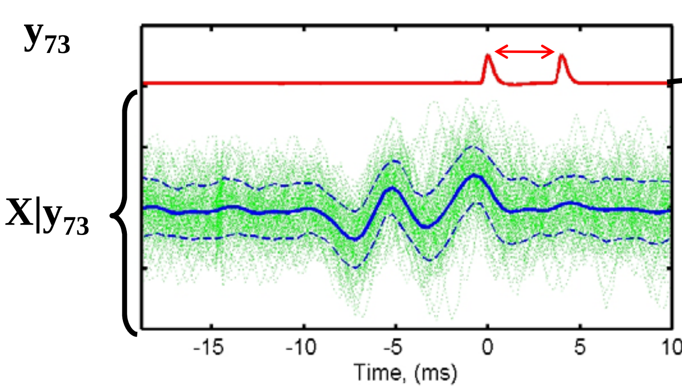
- *$I(X_M;Y_N)$* cannot be estimated directly for high-dimensional stimulus sets – *$P(X,Y)$* not known. Use a model. This produces an upper bound to the distortion (Data processing theorem). **Better model = tighter bound.**
- Here we use a Gaussian estimate of the stimulus:
$$p_{est}(x \mid m) = N(x; \mu_m, C_m)$$
where (μ_m, C_m) are the stimulus mean and covariance, conditioned on class *m.* The estimate of the stimulus probability then is the Gaussian mixture model
$$p_{est}(x) = \sum_{m=1..M} p_m N(x; \mu_m, C_m)$$
The parametric quantizer is $p(m \mid x) = \dfrac{p_m N(x; m_m, C_m)}{p_{est}(x)}$.
- ***!!!* This model imposes a distance on the input space: it defines when stimuli are close to each other.**



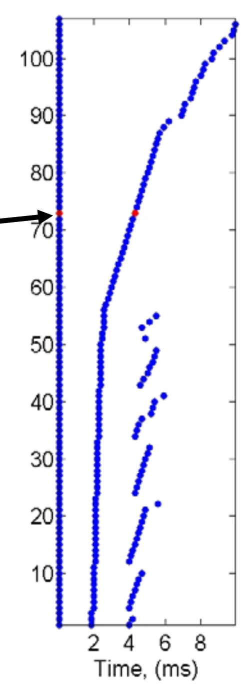
The Cricket Cercal System



Neural Responses (sequences of spikes, T=10 ms) caused by a white noise wind stimulus (Gaussian distribution, 5-500 Hz).

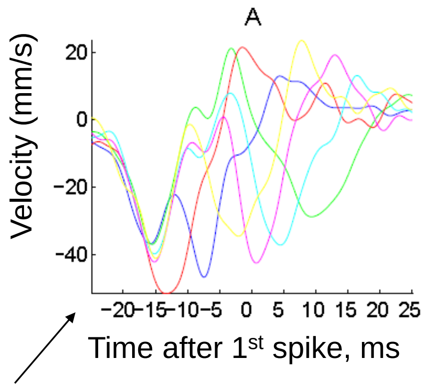


Some of the air current stimuli preceding spike pattern #73 (first spike at Time = 0 ms).

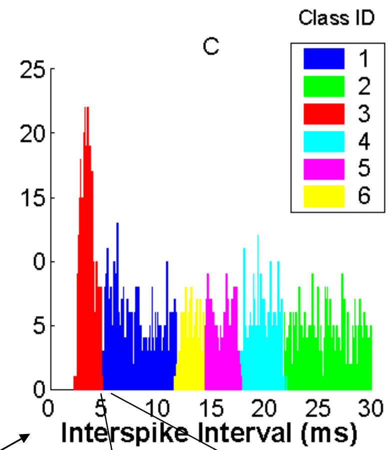


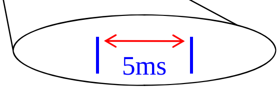
Neural Responses (these are all patterns of length 10 ms

Quantization for inter-spike intervals .



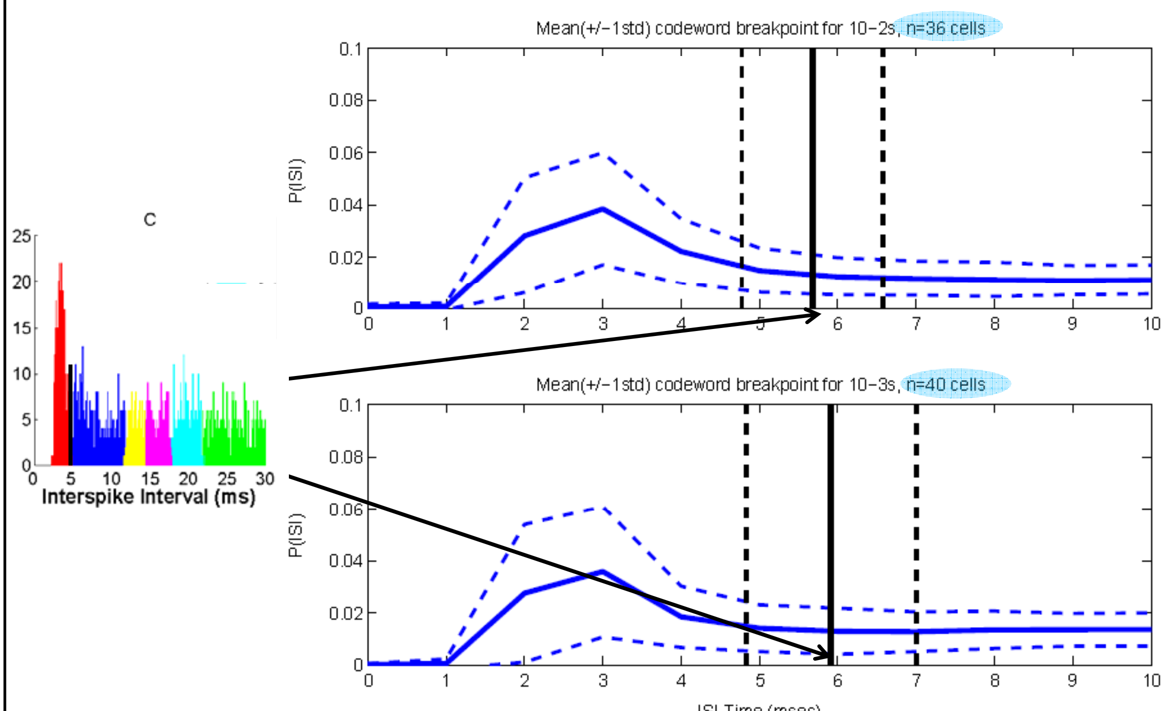
Induced clustering of stimuli





Induced clustering of responses

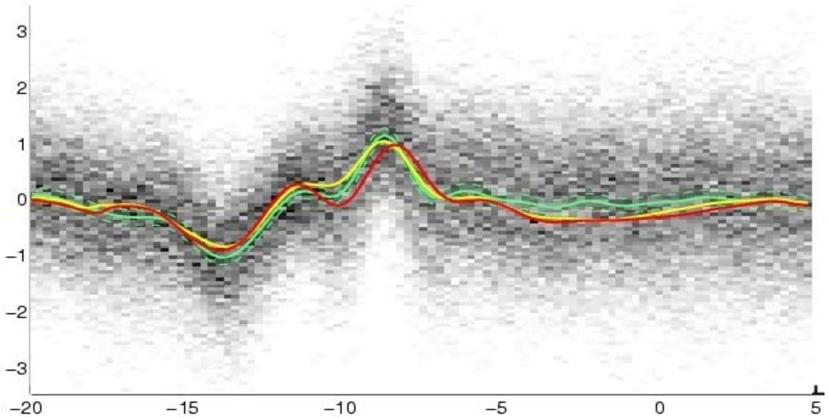
Code features consistent between individuals



Applying the algorithm to cricket sensory data.

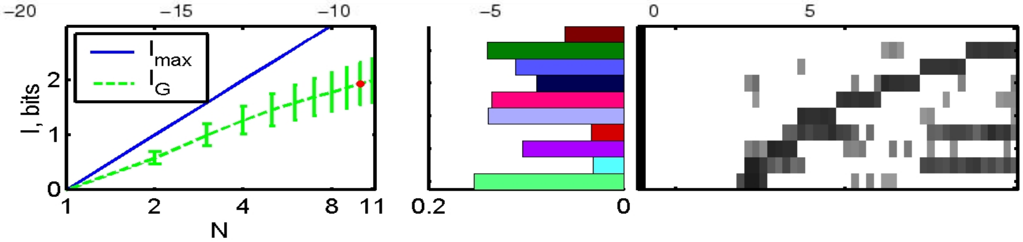
Single cell, unidirectional GWN.

A sequence of refinements in a single cell, along with the class conditioned mean stimuli.

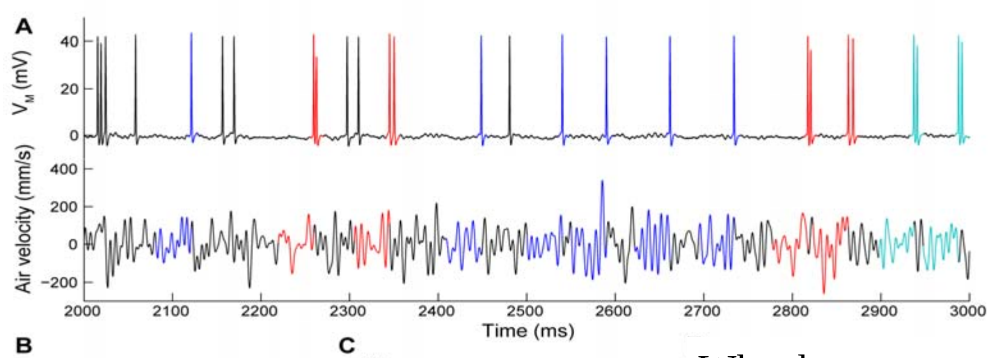


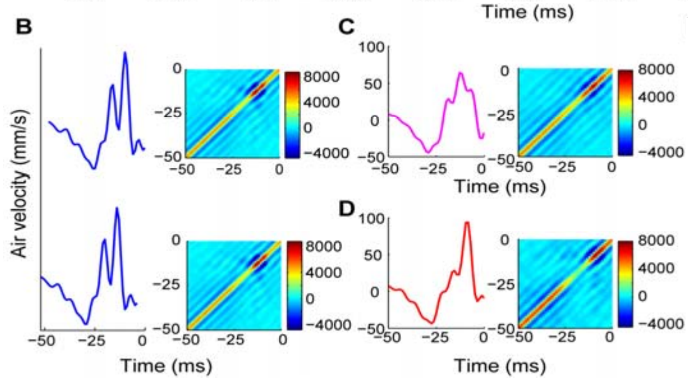
The class conditioned mean (**green**) is superimposed on the stimulus density around spike patterns from **class 10**. The result from linear stimulus reconstruction is shown in **red**. This cell is **not** linear. The yellow trace is the linear stimulus reconstruction of the same class with ISI reduced by 0.4ms.

(Zane, Travis)



Are responses linear?

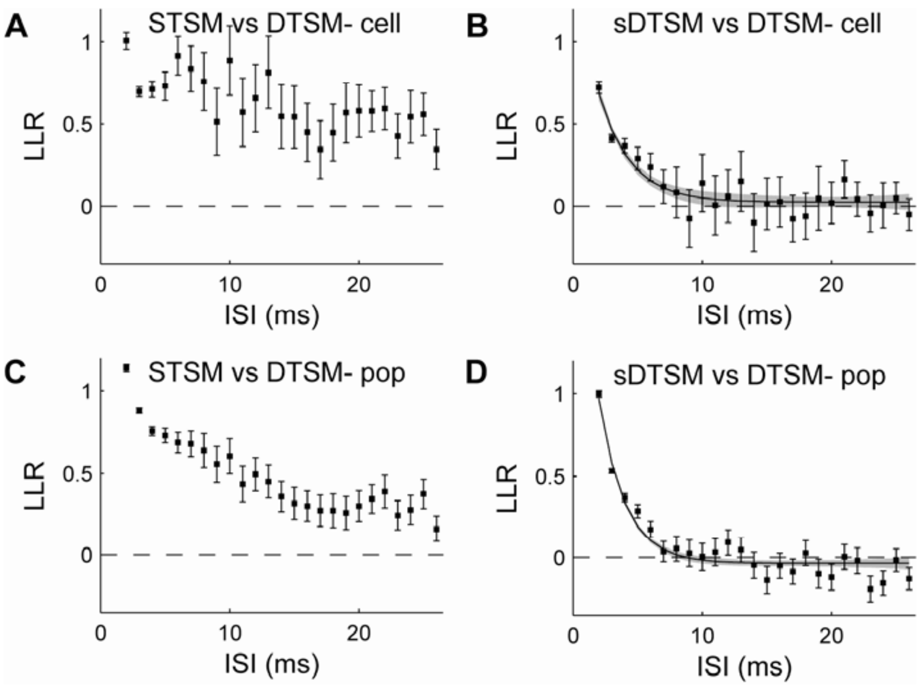




Why do we care: Linear systems are easier to characterize and there is more data for their characterization.

Aldworth et.al. 2010 (under review)

Which responses are linear?

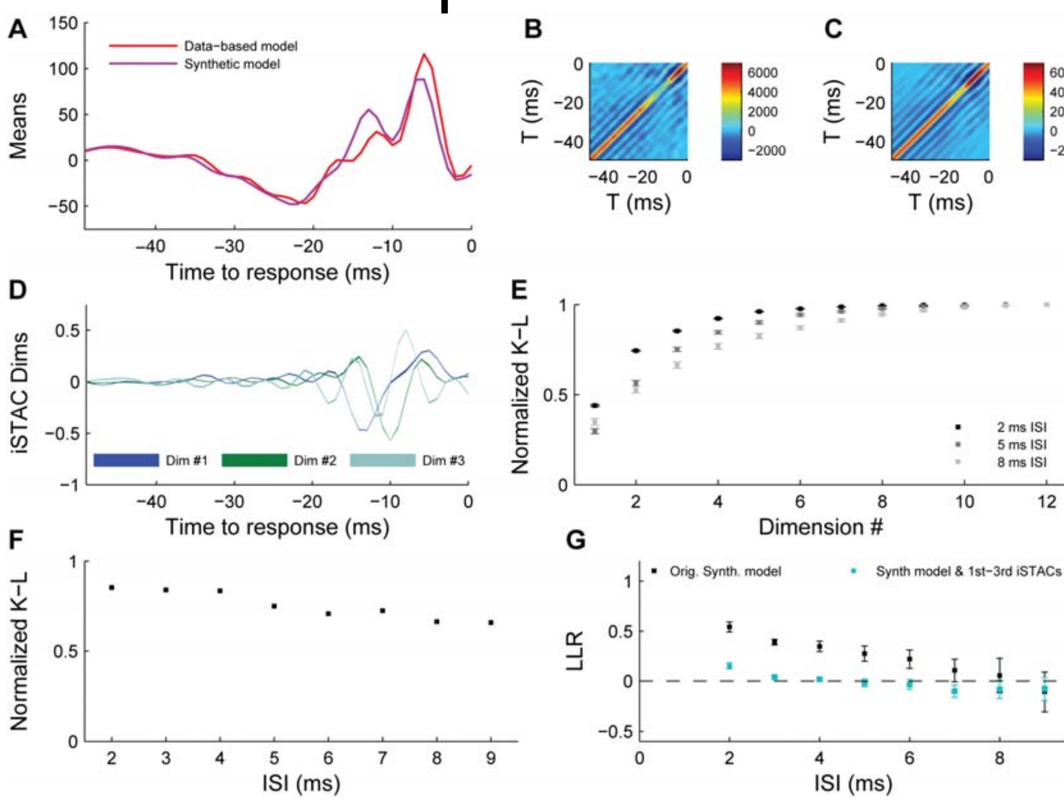


A, C: single vs doublet models

B, D: doublet vs 2-spike linear model

Top: cell; Bottom: population

How are responses non-linear?



Discussion

- model a set of neurons as a communication channel.
- define a coding scheme through equivalence classes of stimulus/response pairs.
    - Coding is probabilistic on codewords.
    - Coding is almost deterministic on codeword classes.
    - The number of classes is $\sim 2^{I(X,Y)}$.
- propose a new method to quantify neural spike trains.
    - Quantize the response patterns to a smaller space.
    - Use an information-based distortion measure.
    - Minimize the information distortion for a fixed size reproduction.
- present results with cricket sensory data.
    - Use temporal patterns of spikes across a few neurons.
    - Recover the stimulus reconstruction kernel at the coarsest quantization.
    - Demonstrate the presence of additional structure at finer quantizations.
    - Demonstrate non-linear processing in several cells.