

Information theoretic approaches

BME665/565

Why Information Theory?

- What forms of neural response are optimal for conveying information about natural stimuli?
- How much does a particular neural response tell us about a stimulus?
- How good is our estimate of the decoding?
- Is there stimulus information in the correlated firings of groups of neurons?

The neuron's view

$$P[s | r]$$



How do we find an optimal s ?

Conditional firing rate

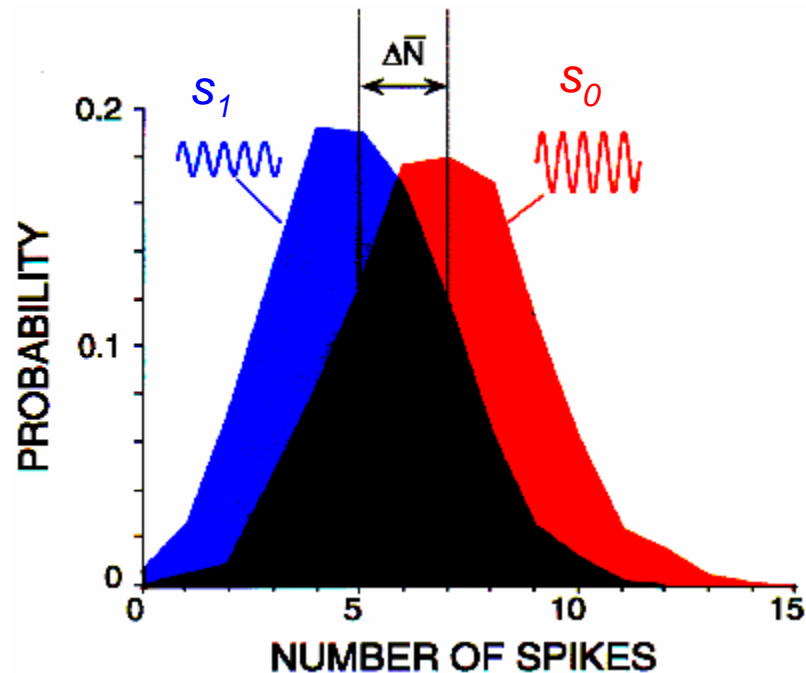
$$P[s | r] = \frac{P[r | s]P[s]}{P[r]}$$

Probability of stimulus (the *prior*)

Probability of response r

Can we use this to discriminate between different stimuli?

Discriminating between possible stimuli



We know that neuronal responses are noisy

We know that populations of neurons may have similar (but not identical) responses to different stimuli

- The Problem:
 - Choose between two stimuli s_0 and s_1 with prior probabilities $P(s_0)$ and $P(s_1)$ given the neural response r so as to minimize the probability of error
 - Conditional firing rate probabilities $P(r | s_0)$ and $P(r | s_1)$ are known

Bayesian Optimal Decision Rule and the Likelihood Ratio

- Bayes' Optimal Decision Rule:
 - Given a specific value of r , choose the alternative which maximizes the posterior probability:

$$\text{if } P[s_1 | r] > P[s_0 | r], \text{ then choose } s_1$$

- Equivalent Likelihood Ratio Test (LRT):

$$\text{if } \frac{P[r | s_1]P[s_1]}{P[r]} > \frac{P[r | s_0]P[s_0]}{P[r]}$$

$$\longrightarrow \text{if } \frac{P[r | s_1]}{P[r | s_0]} > \frac{P[s_0]}{P[s_1]} \text{ then choose } s_1$$

- Likelihood ratio $LR = \frac{P[r | s_1]}{P[r | s_0]}$

Properties of the Likelihood Ratio

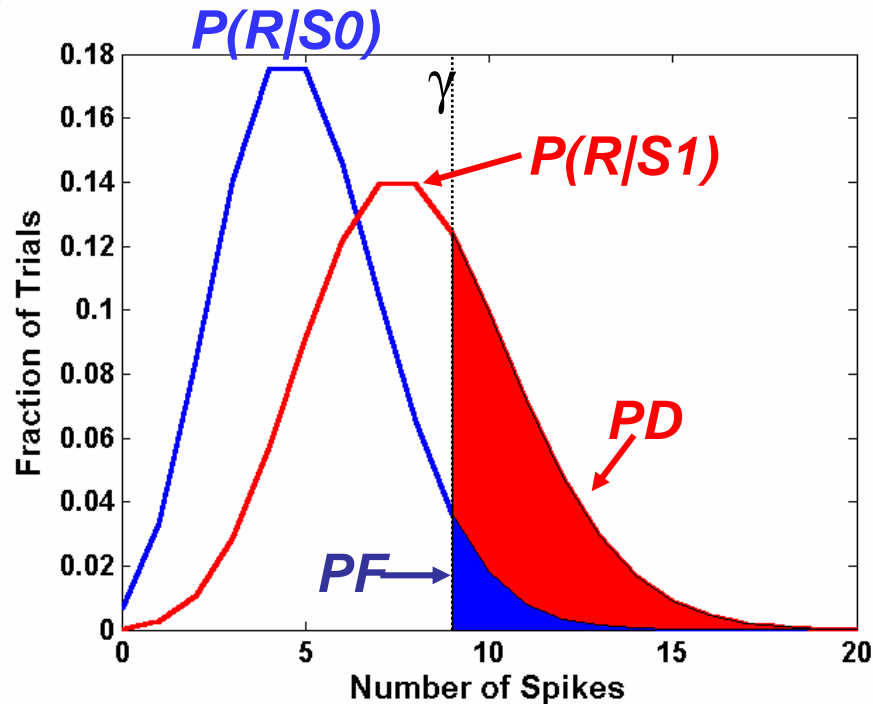
- If the conditional probabilities $P[r | s_i]$ are either Poisson or Gaussian with equal variances, LRT reduces to comparing the mean response difference with a threshold γ :

$$P[r | s] = \frac{(f(s)T)^n}{(n)!} e^{-f(s)T} \quad P[r | s] = e^{-\frac{f(s)}{2\sigma^2}} \quad f(s) = \langle r \rangle$$

$$\text{if } \frac{P[r | s_1]}{P[r | s_0]} > \frac{P[s_0]}{P[s_1]} \text{ then choose } s_1 \implies \text{if } \Delta \langle r \rangle > \gamma \text{ then choose } s_1$$

- LRT works for single neuron responses and for population responses
- Decision rule invariant to monotonic transformation (e.g. logarithm)

Detection and False Alarm Probabilities

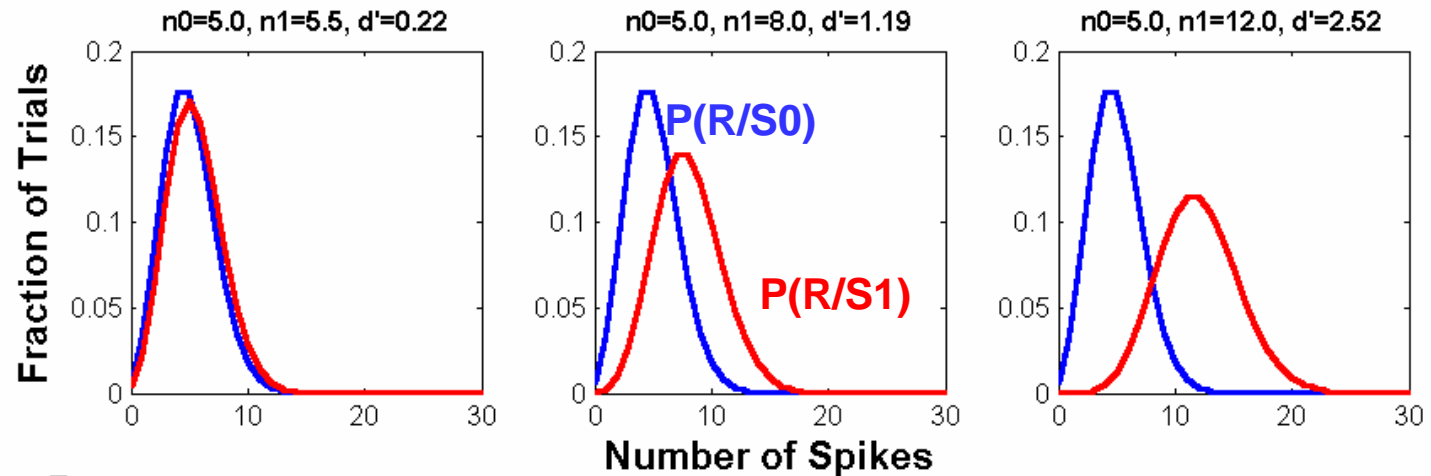


Poisson
Example

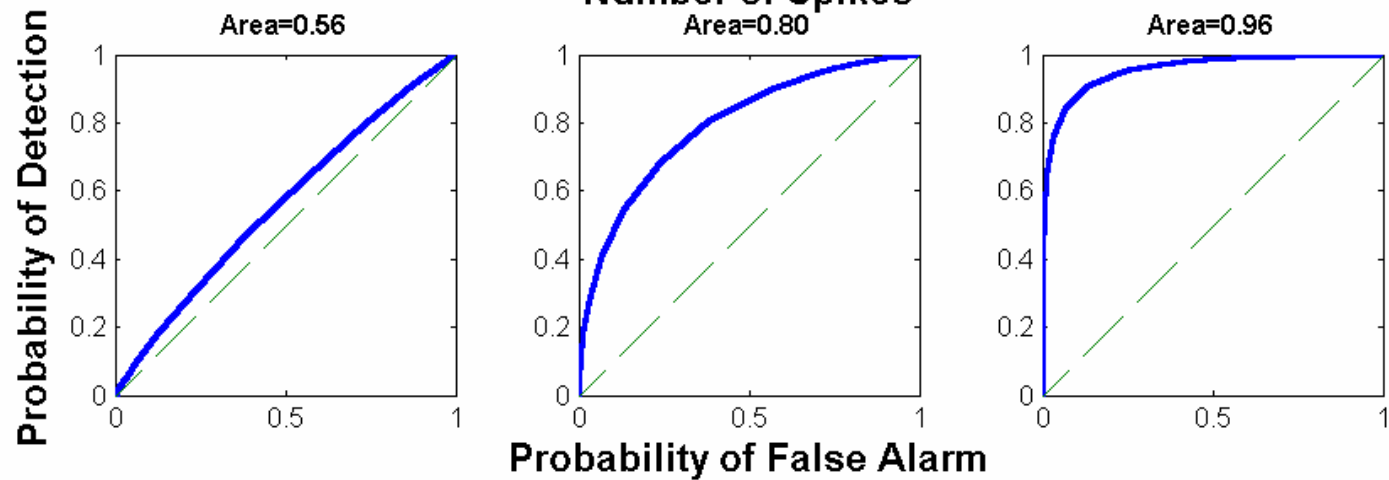
- The neural response (in general LR) is compared to a criterion to make a decision about which stimulus was presented
- Two types of errors: “misses” and “false alarms”
 - PF is the probability of false alarm
 - PD is the probability of detection; $PM = 1 - PD$ is the probability of a miss

Receiver Operating Characteristic (ROC)

Conditional Probabilities (Poisson)



ROC Curves



- The area under the ROC curve gives a distribution-free measure of performance

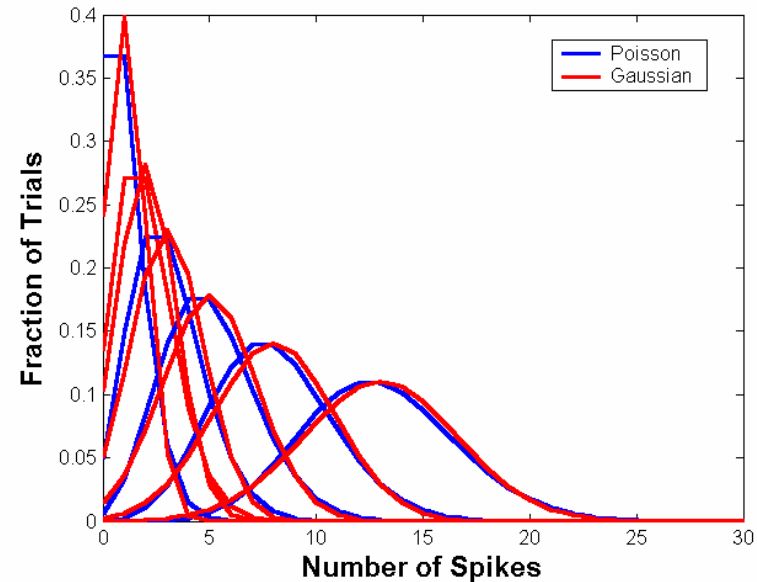
d' as a measure of performance

- If the conditional probabilities $P(r | s_0)$ and $P(r | s_1)$ are Gaussian with equal variances, then

$$d' = \frac{\Delta r}{\sigma}$$

completely determines the performance (the ROC curve)

- If variances are unequal under the two alternatives, they can be averaged
- Many probability distributions (including Poisson) approach a Gaussian when the mean response becomes moderately large
- d' is a poor measure of performance when the number of spikes is very small or when the two stimuli are widely separated



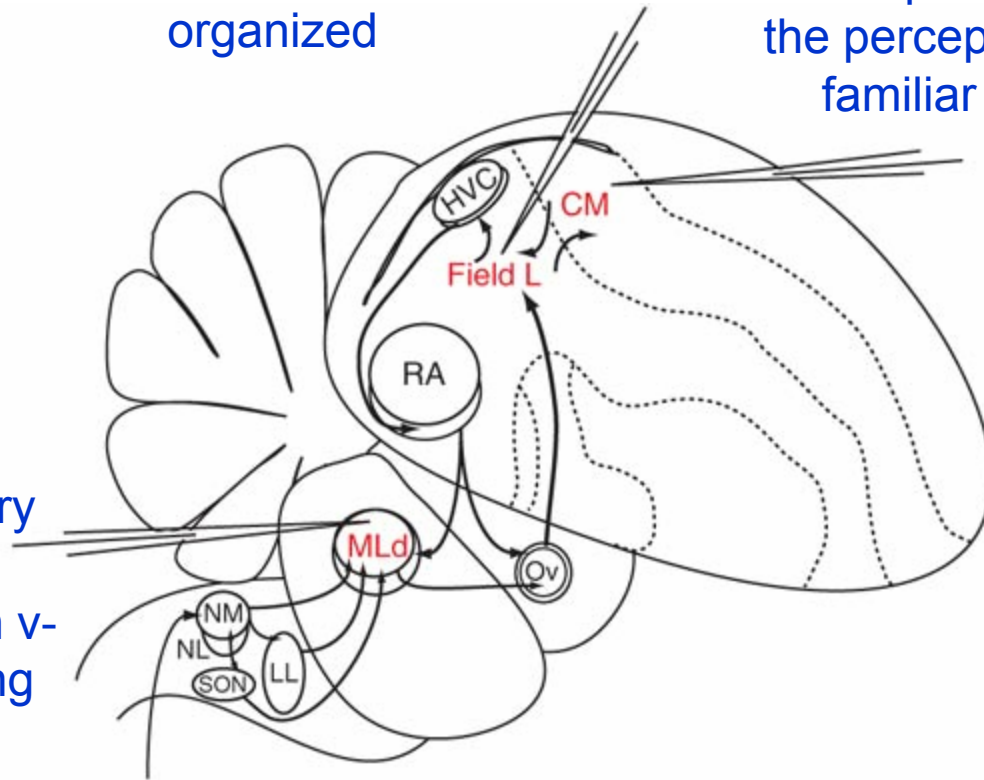
Example: Auditory response to bird song



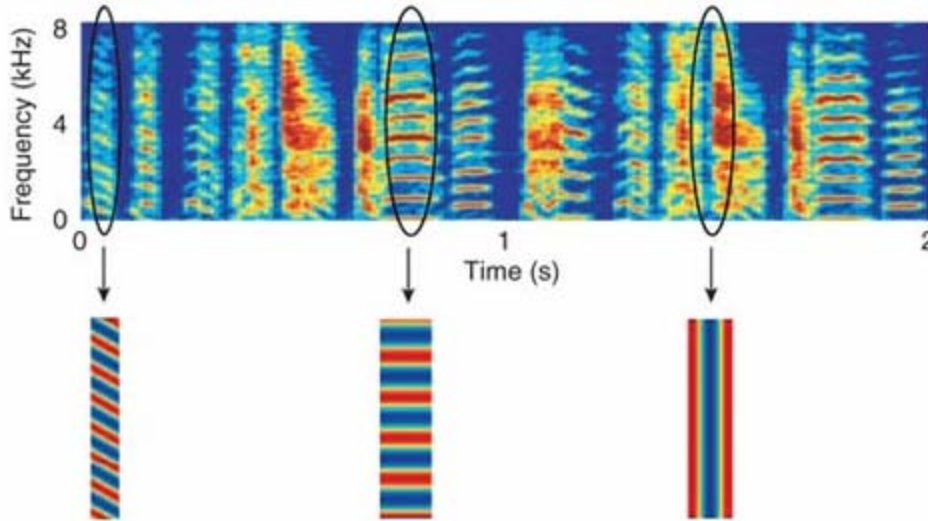
Field L: primary auditory region, tonotopically organized

CM: implicated in the perception of familiar sound

MLd: auditory midbrain, tonotopic with v-shaped tuning curves



Encoding the stimulus characteristics

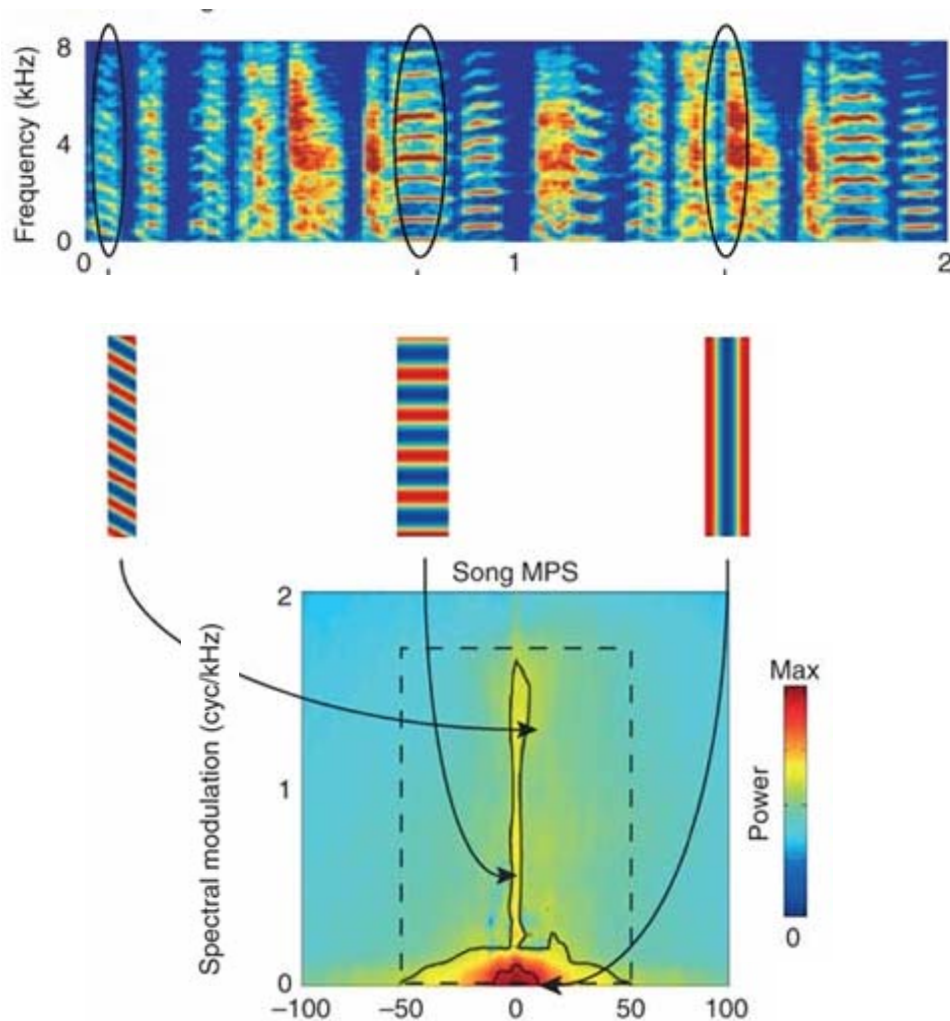


A spectrogram of zebra finch song with example spectro-temporal modulation patterns below. Red indicates high intensity and blue indicates low intensity.

Spectrogram: a windowed discrete-time Fourier transform of a signal using a sliding window.

Modulation: how much a signal varies around a carrier signal.

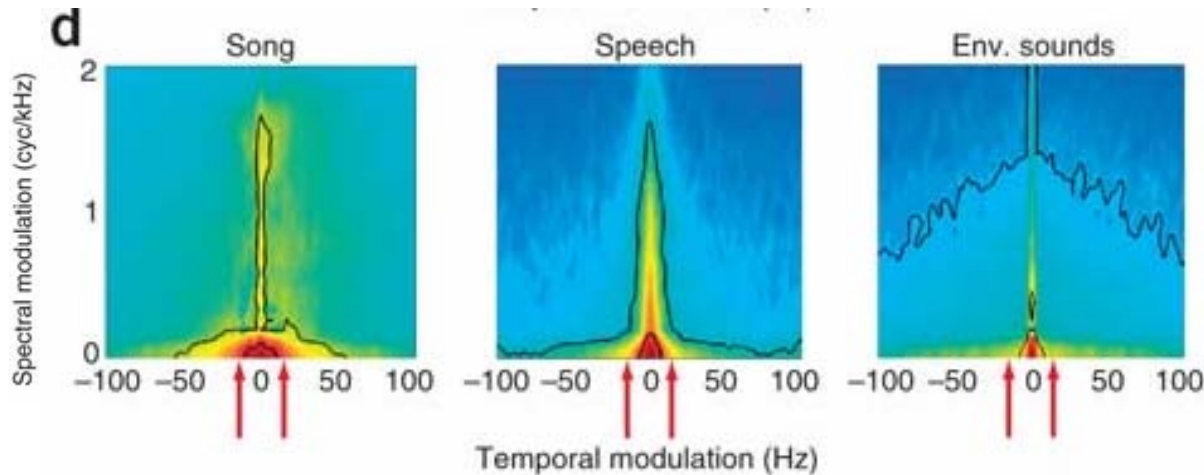
Bird song (stimulus) characteristics



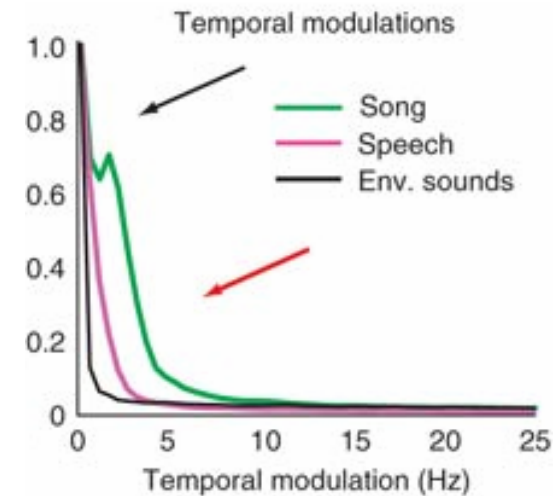
Modulation power spectra (MPS) for song and modulation-limited noise were calculated by taking the two-dimensional Fourier transform of the auto-correlation matrix of the sound spectrogram

The song MPS shows that zebra finch song contains a limited range of modulations; high frequency spectral modulations occur at low temporal modulation frequencies, and high temporal modulations occur at low spectral modulation frequencies

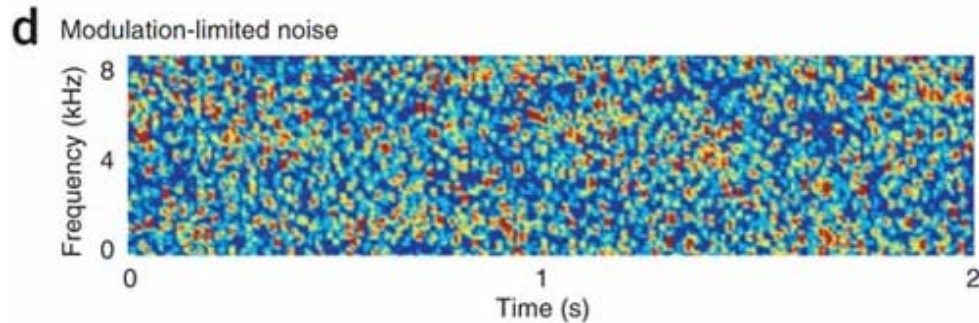
Comparing bird song to other stimuli



- All show high power at low frequencies and rapid decreases in power as frequency increases
- Sound classes differ in the distribution of power across temporal modulation frequencies



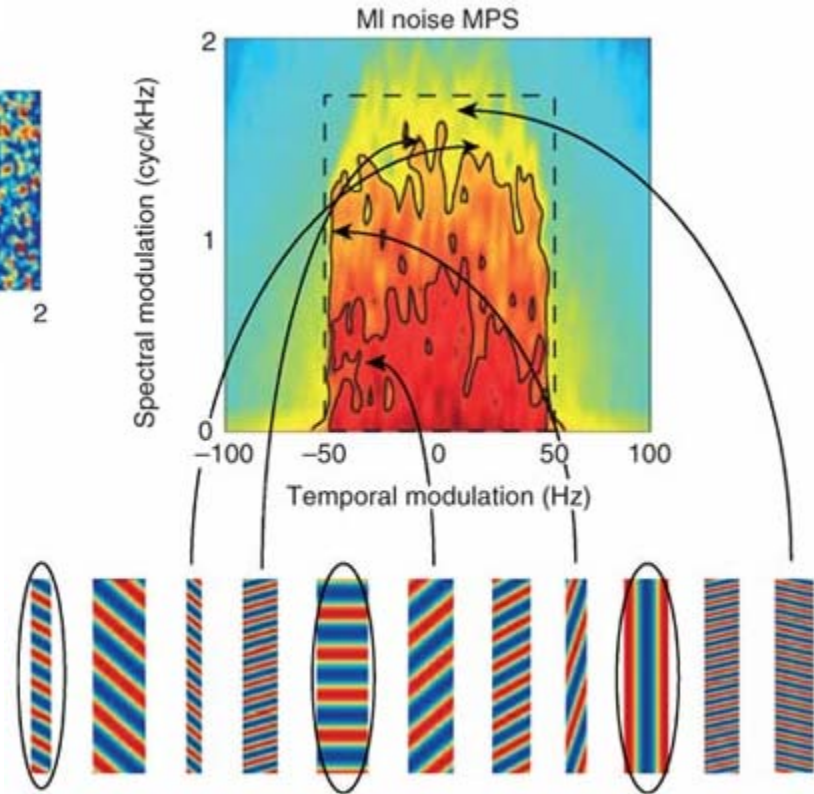
Noise stimulus characteristics



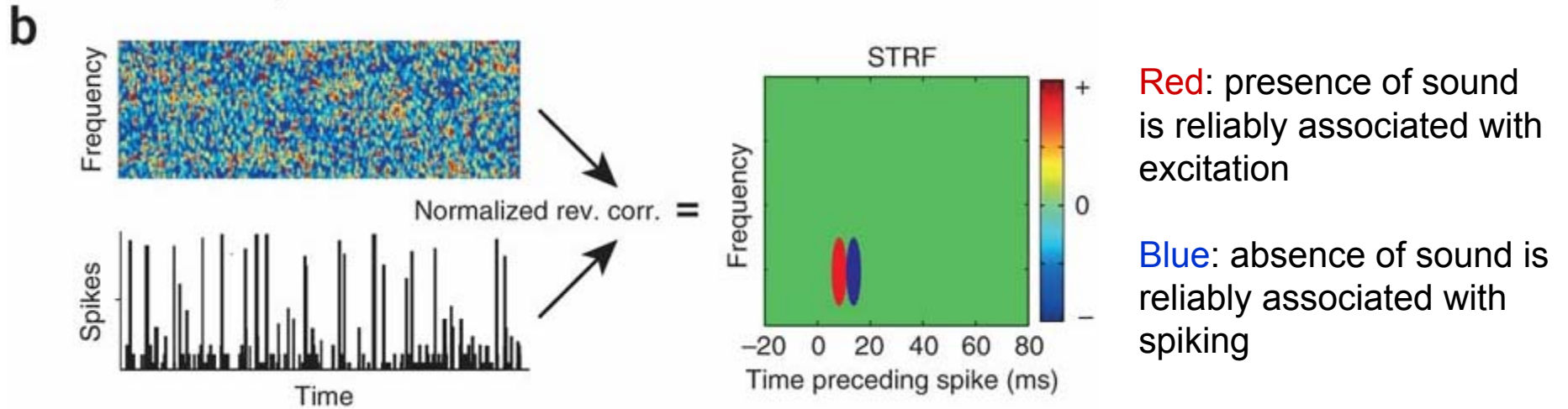
A spectrogram of the modulation-limited noise

For finch song:

- temporal modulations <50 Hz
- spectral modulations <2 cyc/kHz



Spatiotemporal receptive fields



- STRF calculated using a generalized reverse correlation method:

$$r_{est}(t) = r_0 + \int_0^{\infty} D(\tau) s(t - \tau) d\tau$$

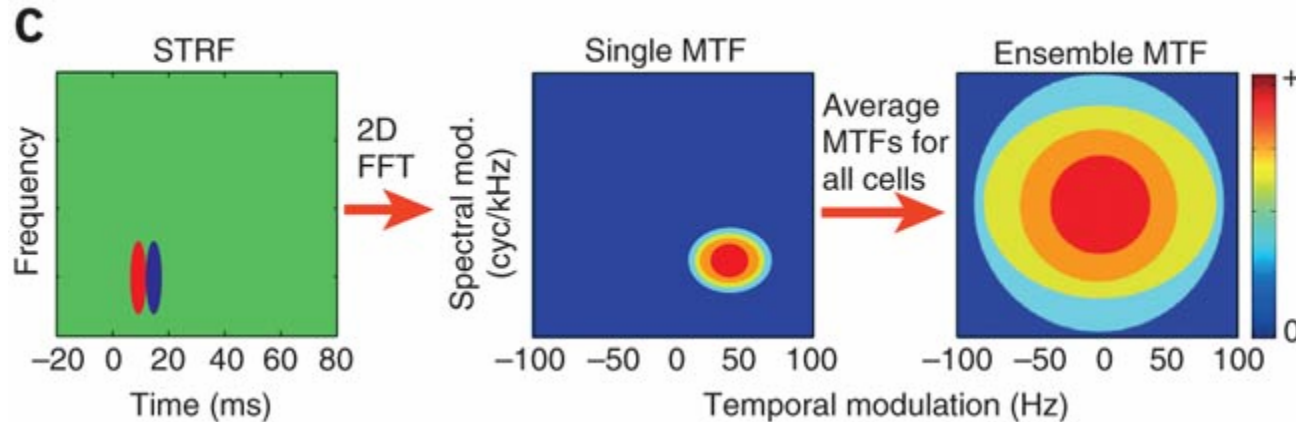
$$\int_0^{\infty} Q_{ss}(\tau - \tau') D(\tau') d\tau' = Q_{rs}(-\tau)$$

$$\hat{r}[t] = \sum_{i=0}^{N-1} \sum_{k=0}^{M-1} h_t[i, k] s[t - i, k]$$

Minimize: $\langle (\hat{r} - r)^2 \rangle$

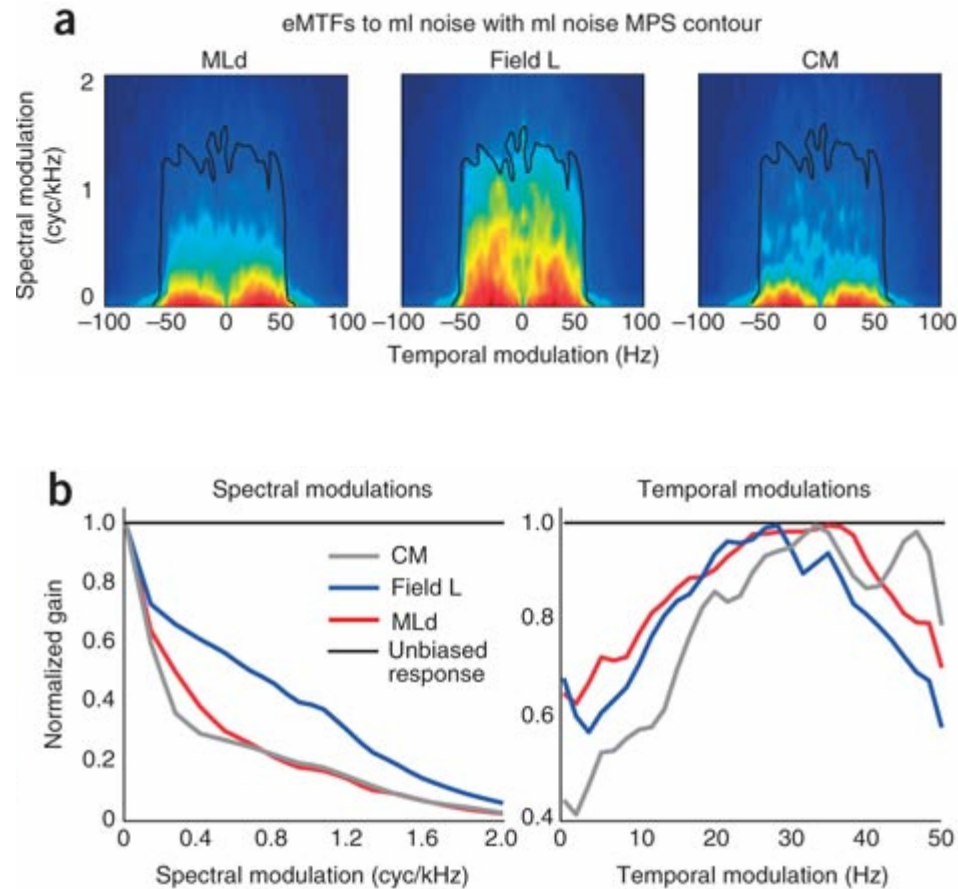
$$h = C_{ss}^{-1} C_{sr}$$

Modulation transfer function: tuning of the neuron



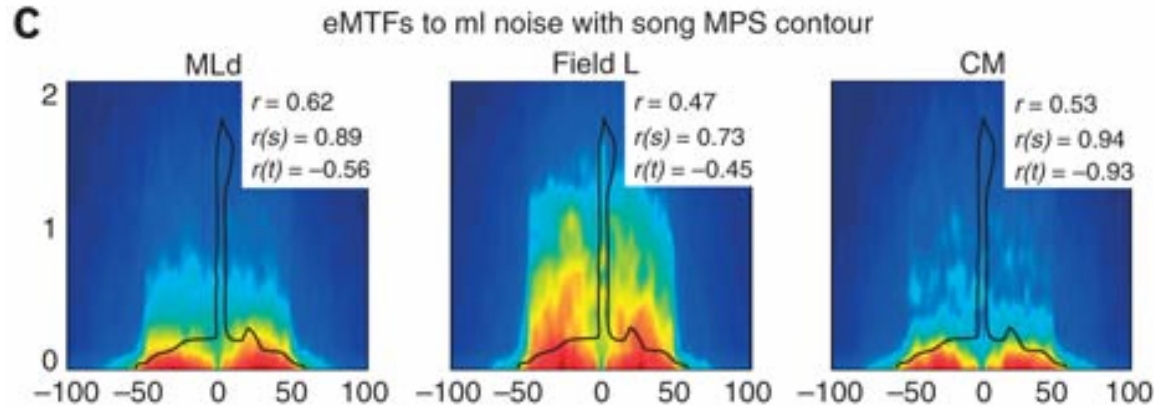
- MTF is the *spatial* frequency response of a signal; it is the contrast at a given spatial frequency relative to low frequencies.
- The modulation transfer function (MTF) for a neuron was calculated by taking the modulus of the two-dimensional Fourier transform of the STRF
- Ensemble modulation tuning for all the neurons in one brain region was measured by averaging the MTFs for all cells to get a single ensemble modulation transfer function (eMTF)
- Allows comparison between modulation characterizing neuronal response and modulations characterizing types of stimuli

eMTFs: Response to modulated noise



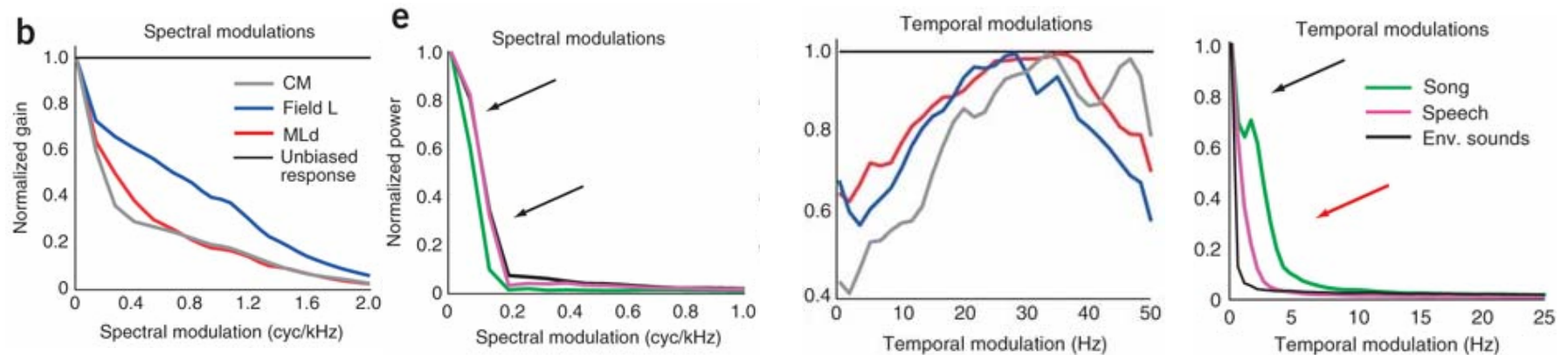
- Spectral modulation tuning low pass in all three auditory regions
- Temporal modulation tuning was band-pass in all regions
- Neurons in Field L have broader spectral tuning curves (perhaps because there are more types of neurons)

eMTFs : Similarity to bird song

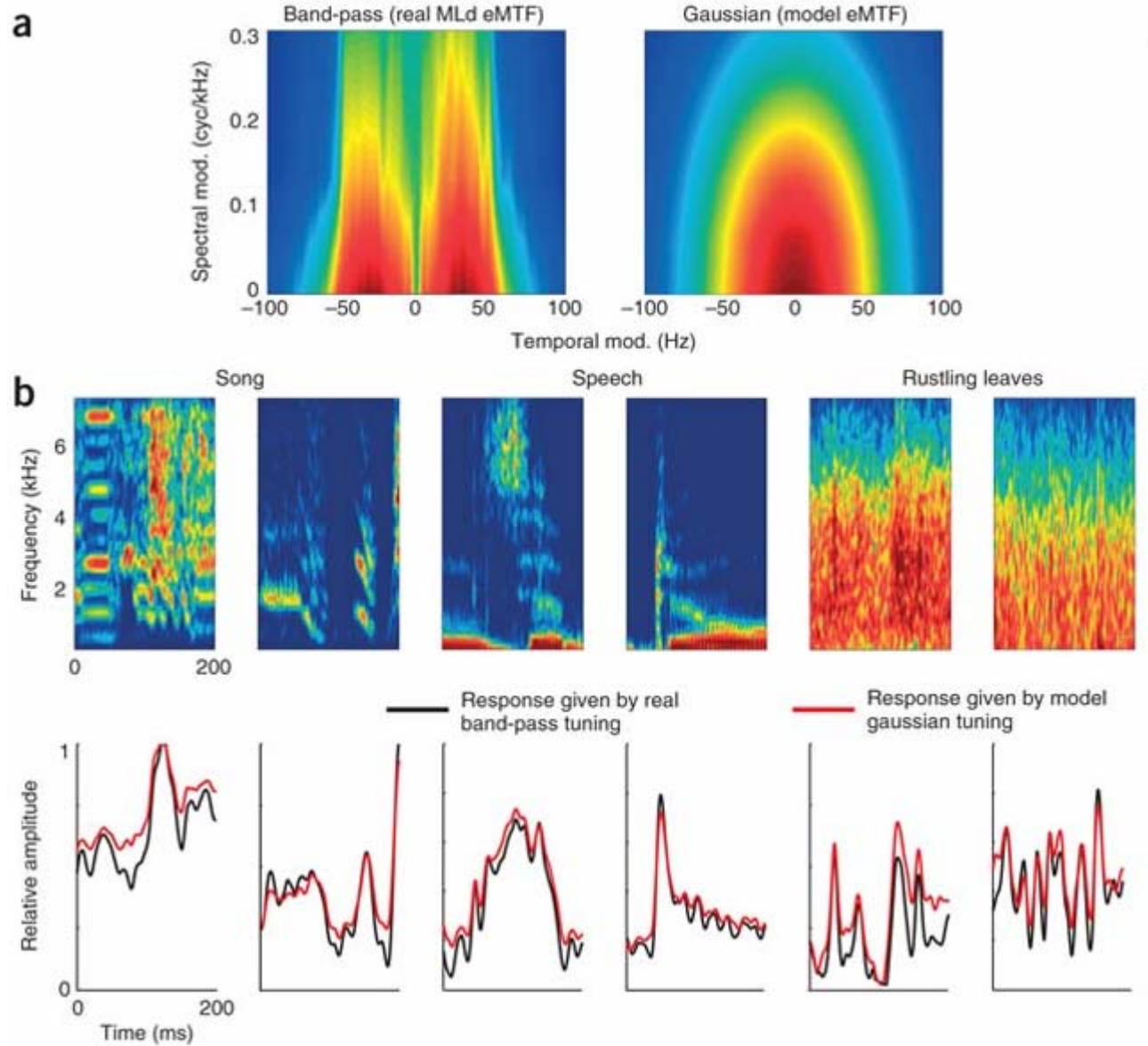


The low-pass spectral modulation tuning matches zebra finch song in that most of the power in song occurs at low spectral modulations

- In contrast, the modulations to which the most tuning gain was devoted were not the most strongly represented temporal modulations in song.



Comparison to a Gaussian tuning curve

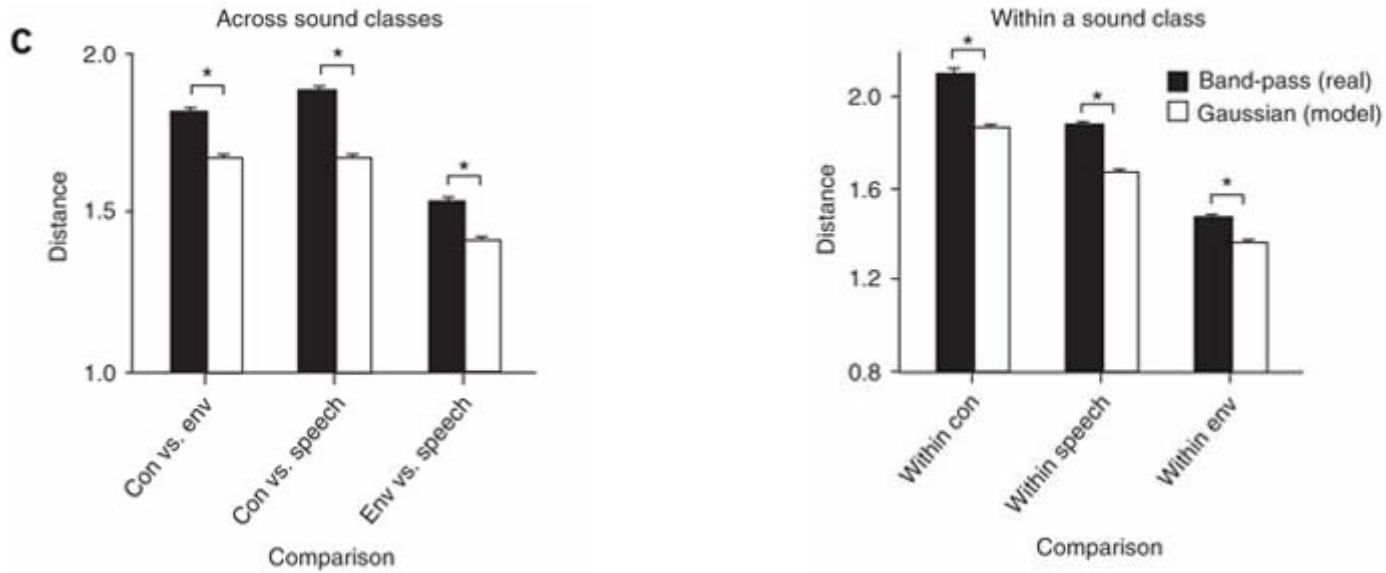


eMTF's of MLd and Gaussian model

Spectrograms of sample stimuli

Predicted ensemble responses (convolution of stimulus and STRF obtained from the eMTF's)

Performance of the Gaussian Model



D' uses the Euclidean distance between the time-varying responses to two sound segments:

$$d' = \frac{\Delta r}{\sigma} \quad \left. \vphantom{d'} \right\} \rightarrow D' = 2 \cdot \sqrt{\sum_{t=1}^{N_{\text{TRin}}} \frac{(R_t(A) - R_t(B))^2}{\sigma_c^2}}$$

Entropy rate of spike trains

- The entropy rate of a spike train gives an upper bound on the information about the stimulus that can be transmitted by the spike train
- It roughly measures how “surprising” a set of responses is – the amount of information about the stimulus carried by the response
- We define a measure of surprise associated with the probability of a response r , as: $h(P[r])$
- We assume responses from two neurons are independent, and require that $h(P[r_1]P[r_2]) = h(P[r_1]) + h(P[r_2])$ which leads us to use a logarithm for h

- By convention, the logarithm is base 2, and the information is reported in “bits”:

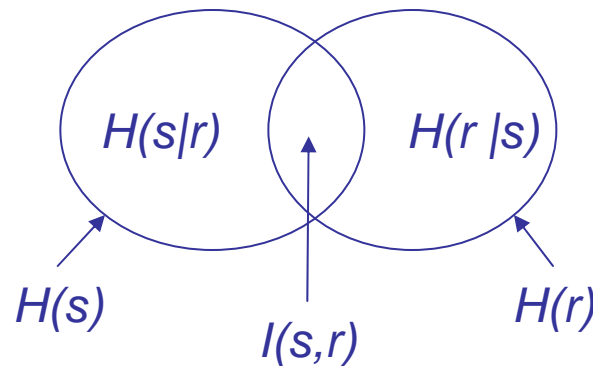
$$h(P[r]) = -\log_2 P[r]$$

- Then, the Shannon entropy measure is the average of h over all possible responses

$$H = -\sum_r P[r] \log_2 P[r]$$

Mutual Information

- The entropy $H(s)$ represents the uncertainty about the stimulus in the absence of any other information
- The conditional entropy $H(s|r)$ represents the remaining stimulus uncertainty after the neural response has been measured
- $I(s,r)=H(s)-H(s|r)$ is the mutual information between s and r ; it represents the reduction in uncertainty achieved by measuring r



Calculating I_m

- $H(s|r)$ is hard to measure because it requires estimating the stimulus from neural response
- By symmetry, $I(s,r) = H(r) - H(r|s)$
- $H(r|s)$ is the entropy of the part of the neural response that is NOT predictable from the stimulus, i.e. the noise in the response

$$H_s = -\sum_r P[r|s] \log_2 P[r|s]$$

- If we average over repetitive presentation of the same stimulus, we get an estimate of the entropy of the noise:

$$H_{noise} = \sum_s P[s] H_s = -\sum_r P[s] P[r|s] \log_2 P[r|s]$$

$$I_m = H - H_{noise}$$

$$\longrightarrow I_m = \sum_{s,r} P[s] P[r] \log_2 \frac{P[r|s]}{P[r]}$$

Information rates and coding efficiencies in sensory neurons

Neural system and species	Information rate (bits/spike)	Information rate (bits/s)	Coding efficiency ^a	Best timing precision and variance:mean
Constant stimulus^b				
Cat retinal ganglion cells	~0.04–0.10	0.4–0.8 ^c	–	–
Primary visual cortex (VI) of rhesus monkeys	–	0.62 ^d	–	–
Middle temporal area (MT) of rhesus monkeys	0.025	0.89 ± 0.29	–	~1.3
Inferior temporal area (IT) of rhesus monkeys	~0.13	0.9 ^e	–	–
Hippocampus of rhesus monkeys	~0.18 ^e	0.32 (maximum = 1.2) ^f	–	–
Variable stimulus: reconstruction method				
HI motion-sensitive neuron of a fly	0.75	64	30%	–
Frog auditory afferents	0.66	23	11%	~2 ms
Vibratory receptors of the bullfrog sacculus	2.6	155	50–60%	–
Cricket mechanoreceptors	0.6–3.2	75–294	50–60%	~0.4 ms
Salamander retinal ganglion cells	1.9	3.7 (up to 10 for a population of >10 cells)	26% (>79% for >10 cells)	–
The MT of anesthetized rhesus monkeys	~0.65	6.7 (maximum = 12.3)	–	2–4 ms
The MT of alert rhesus monkeys	0.6	5.5	<30%	–
Variable stimulus: direct method				
HI motion-sensitive neuron of a fly	2.43	80	50%	1.5–3 ms; <0.1
Salamander and rabbit retinal ganglion cells	3.7	16.3	59%	≥0.70 ms; ≥0.05
The MT of alert rhesus monkeys	1.5	12 (maximum = 29)	Up to 45%	<2 ms; ~1.4

Information Theory Pro's and Con's

- Pros
 - Does not assume any particular neural code
 - Can be used to identify the stimulus features best encoded by neurons or to compare effectiveness of different putative neural codes
 - One number summarizes how well stimulus set is coded in neural response
- Cons
 - Information estimate depends on stimulus set. Stimulus probabilities in environment hard to specify.
 - Does not specify how to read out the code: the code might be unreadable by the rest of the nervous system
 - For all but simplest examples, estimation of mutual information requires huge amount of data. Methods that try to circumvent data limitations (e.g. stimulus reconstruction) make additional assumptions (e.g. linearity or Gaussian distribution) that are not always valid.