Information and redundancy in the auditory system

Gal Chechik

Current address: Robotics lab, CS department, Stanford

http://robotics.stanford.edu/~gal

Joint work with

Israel Nelken	Hebrew University
Omer Bar-Yosef	Hebrew University
Mike Anderson	Johns Hopkins
Eric Young	Johns Hopkins
Jan Schnupp	Oxford
Andrew King	Oxford
Naftali Tishby	Hebrew University

Preview

- An empirical comparison of methods of mutual information (MI) estimation from spike trains
- In our data, information is well extracted using two simple statistics of spike trains: mean activity time and spike count.
- Using reliable MI estimators we observe changes in coding along the hierarchical auditory pathway.

Outline

- **General concerns in MI estimation**
- Data and methods compared
- Validation of methods on simulated data
- Results with real data
- Simple statistics of spike trains
- Redundancy in the auditory pathway

Information in neural activity

The general setting:

- Given two high dimensional signals: stimuli S and responses R, we wish to measure their relation, as quantified by the mutual information (MI) in their joint distribution I[p(S;R)].

Naïve approach:

- Estimate the density p(S;R), then the MI, I[p].

Improvement:

- Improved density estimation
- Estimate MI/entropy directly (without estimating p)
- Focus on low order statistics of spike trains. p(S;f(R))
  - Added value: hints the for readout mechanism

Choosing statistics of spike trains

The goal: Project R to low dimension without losing MI

When distributions are known (limit of infinite data)

we have $I[p(S,R)] \geq I[p(S,f(R))]$

- We could search for a simple statistic f which maximizes $\max_f I[p(S,f(R))]$

With finite samples the estimators are biased

- Maximization is no longer allowed since it is no longer true that. $I[p(S,R)] \geq \hat{I}[\hat{p}(S,f(R))]$
- We risk to overestimate the MI
- We therefore must control for their bias and variance

Relation to over fitting of classifiers

Evaluation of MI estimators

We wish to estimate MI on known distributions.

These must be similar to the real neurophysiological distributions

- Fit a parametric model to the experimental data.
- "true" information of the model can be calculated
- Use it to validate any MI estimation and bias correction procedures.

Inhomogeneous Poisson is a reasonable model of our data

- Each time point is distributed near Poisson
- Correlation in the data are not large

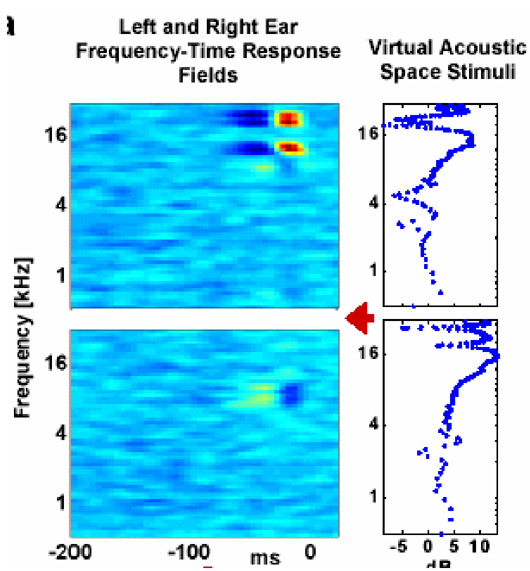
Outline

- General concerns in MI estimation
- ➢ Data and methods compared
- Validation of methods on simulated data
- Results with real data
- Simple statistics of spike trains
- Redundancy in the auditory pathway

Two different data sets

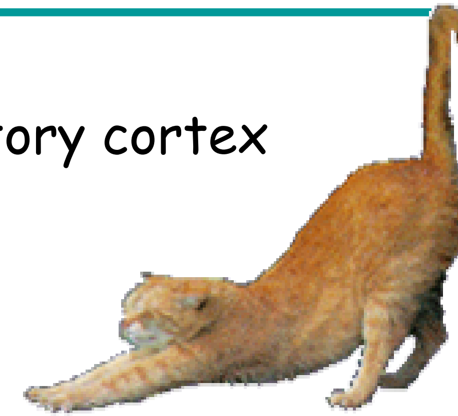
- Ferret auditory cortex

- Virtual space stimuli

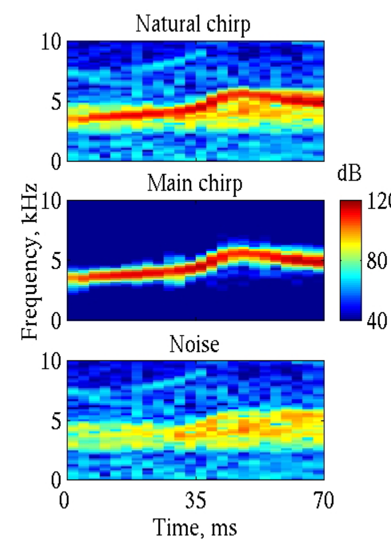


- Deep anesthesia (barbiturate)

- Cat auditory cortex



- Bird chirps



- Light anesthesia (halothane)

MI estimation methods

- Matrix based methods
  - Spike counts
  - Latency of the first spike
  - Mean arrival time of all spikes
  - Spike patterns (the direct method)[Strong et al]
- Binless estimation [Victor]
  - Use nearest neighbors instead of binning
  - Project spike counts to Euclidian space
- 2nd order Taylor expansion [Panzeri & treves]
  - Contains rates, inter and intra train correlations
- ML Decoding algorithm [Treves]
  - Fit the data with a parametric model (Gaussian/Poisson)
  - Classify responses to stimuli using ML (leave one- out)
  - Calculate MI of confusions matrix

Reducing the bias in matrix MI estimators?

- **The naïve MI estimator:** Calculate MI of the empirical distribution

$$\hat{p}(s,r)=\frac{\#\{s,r\}}{\#total};\qquad I[\hat{p}(S,f(R))]=\sum_{s,r}\hat{p}(s,r)\log\frac{\hat{p}(s,r)}{\hat{p}(s)\hat{p}(r)}$$

- **This estimator is biased**
  - To a first order bias ≈ $\frac{\#\text{ degrees of freedom}}{2N\log(2)}$
  - Both for I=0 and I>0 [MM, PT]
  - Bias is distribution dependent (often correlated to MI)
  - Depends on the number of "effective bins"
- **To reduce the bias we apply adaptive binning to achieve near uniform marginals**
  - Iteratively unite rows/columns to approach uniform margin
  - Choose matrix dimension to maximizes PT-bias-corrected MI

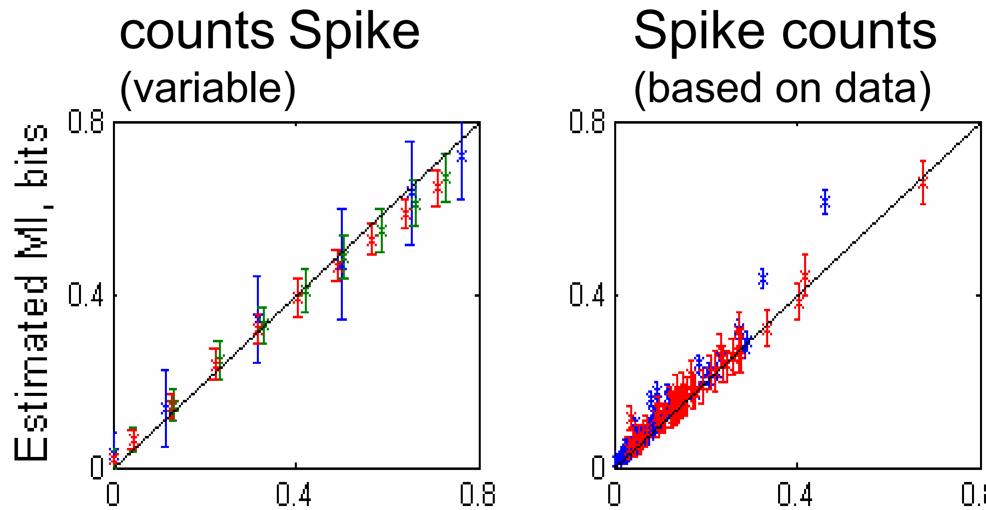
Validating the adaptive binning procedure

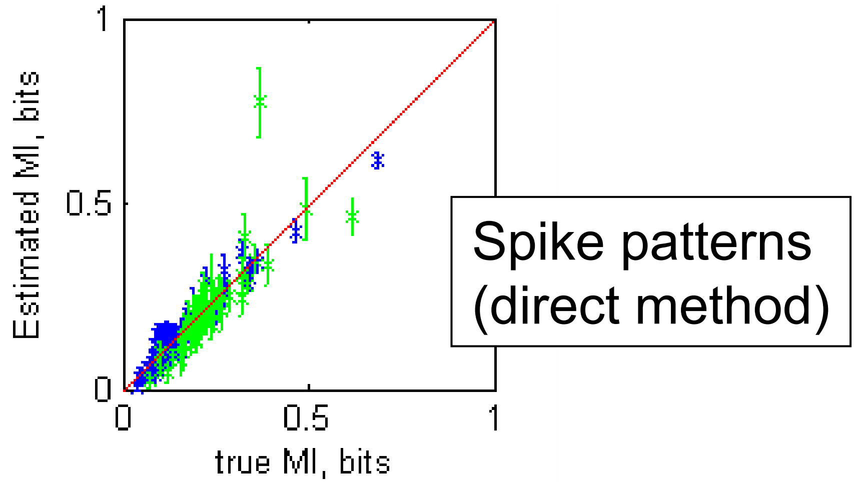
Compare estimated and "true" MI of various statistics:

- Spike counts
- Spike patterns

Models generated both to

- Cover range of parameters
- Based on real data statistics





Error bars: MI std over 10 repeats.

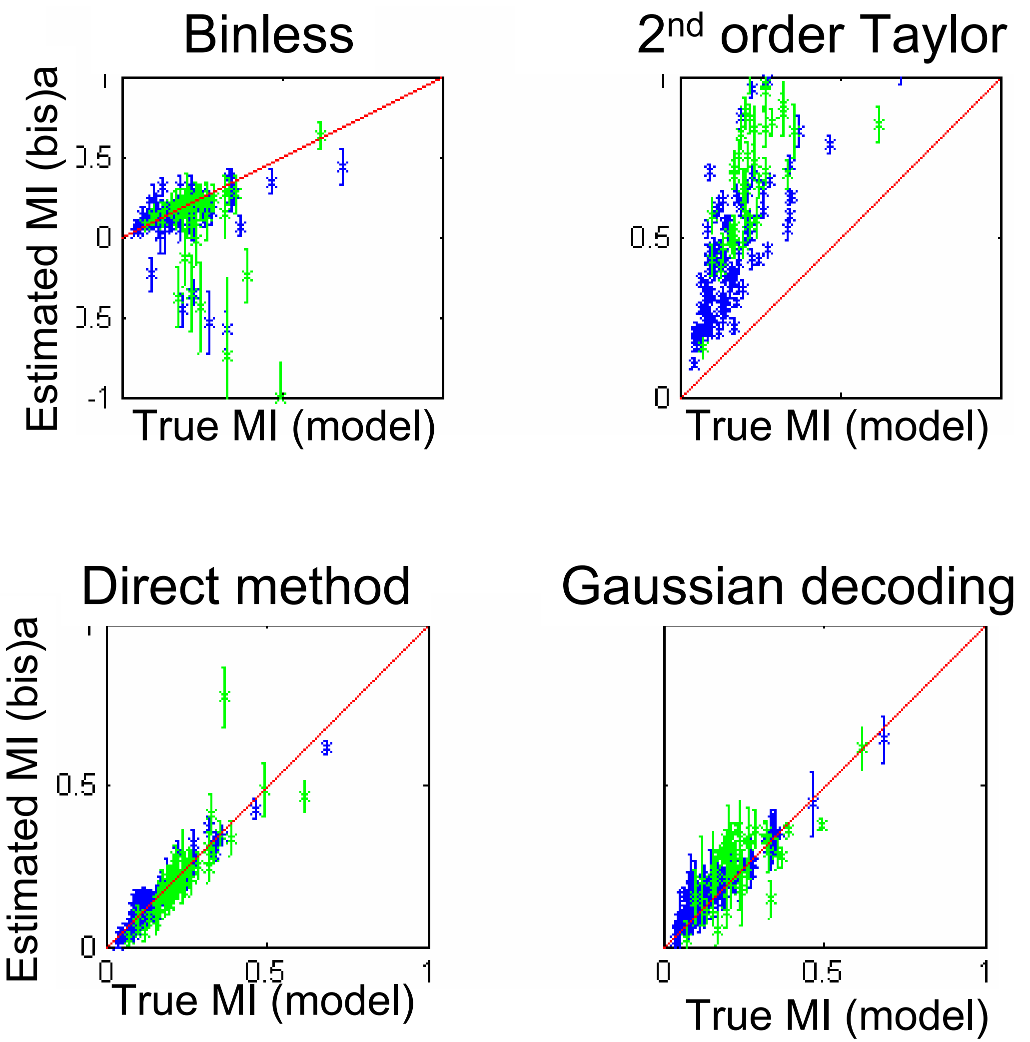
Outline

- General concerns in MI estimation
- Data and methods compared
- ➢ **Validation of methods on simulated data**
- Results with real data
- Simple statistics of spike trains

Comparing MI in simulated data

Compare estimated and "true" MI with various methods

- Binless:
  - MI underestimation
- Taylor expansion
  - MI overestimation
- Spike patterns
- Gaussian decoding



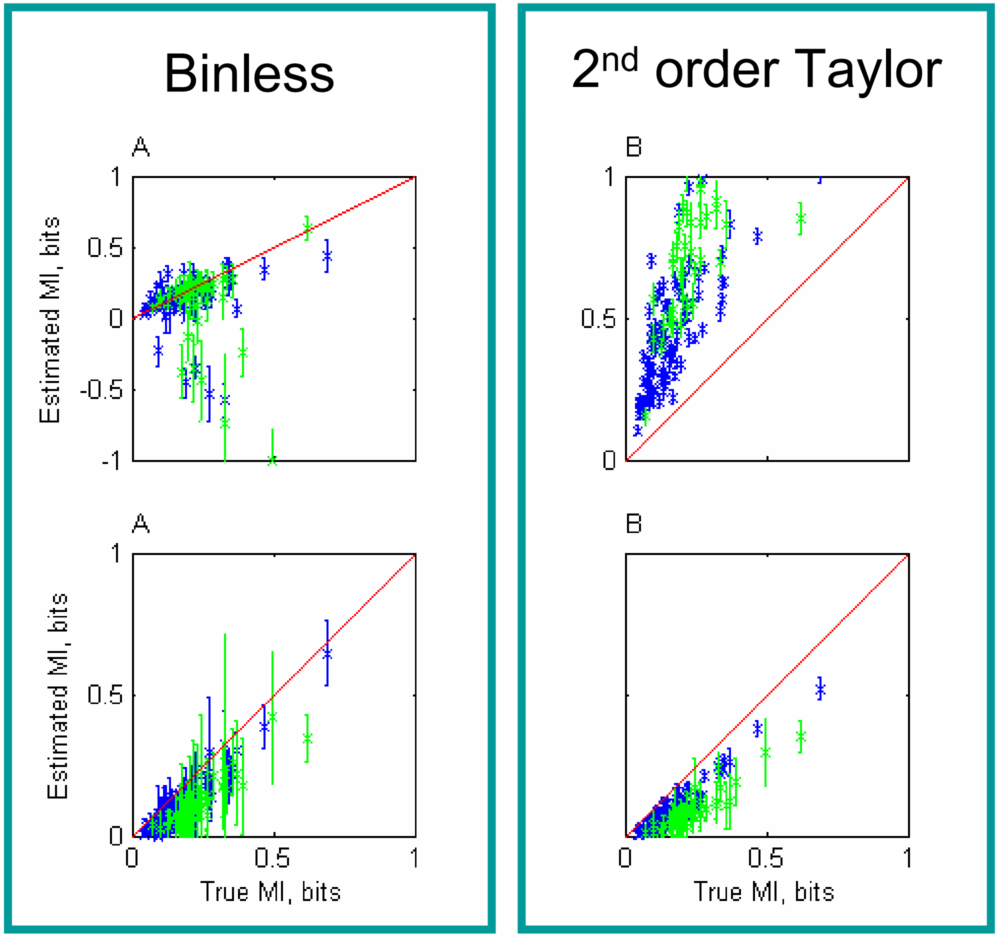
Green = ferrets, Blue = cats

Bias correction using shuffling

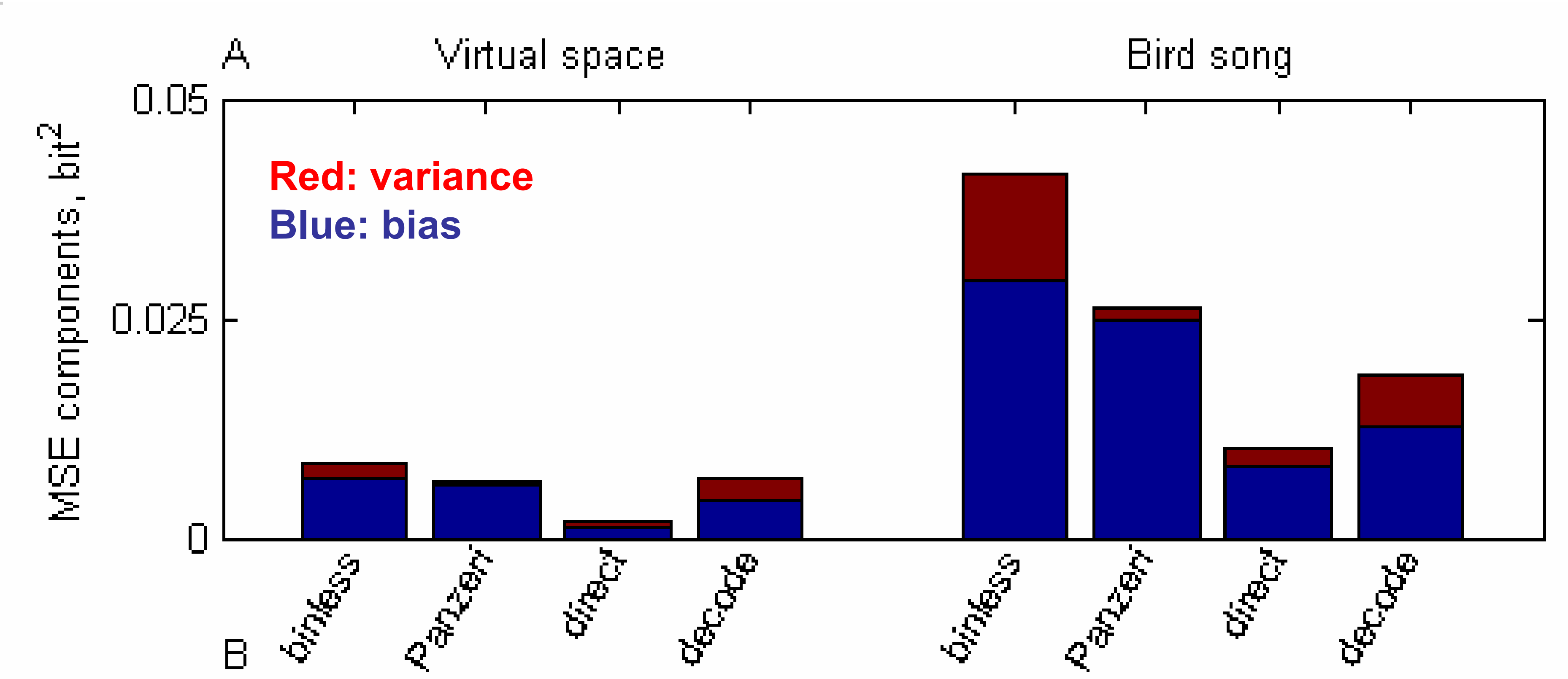
- Using shuffling to correct bias – is not good enough

Without bias correction

With bias correction



Summary of validation

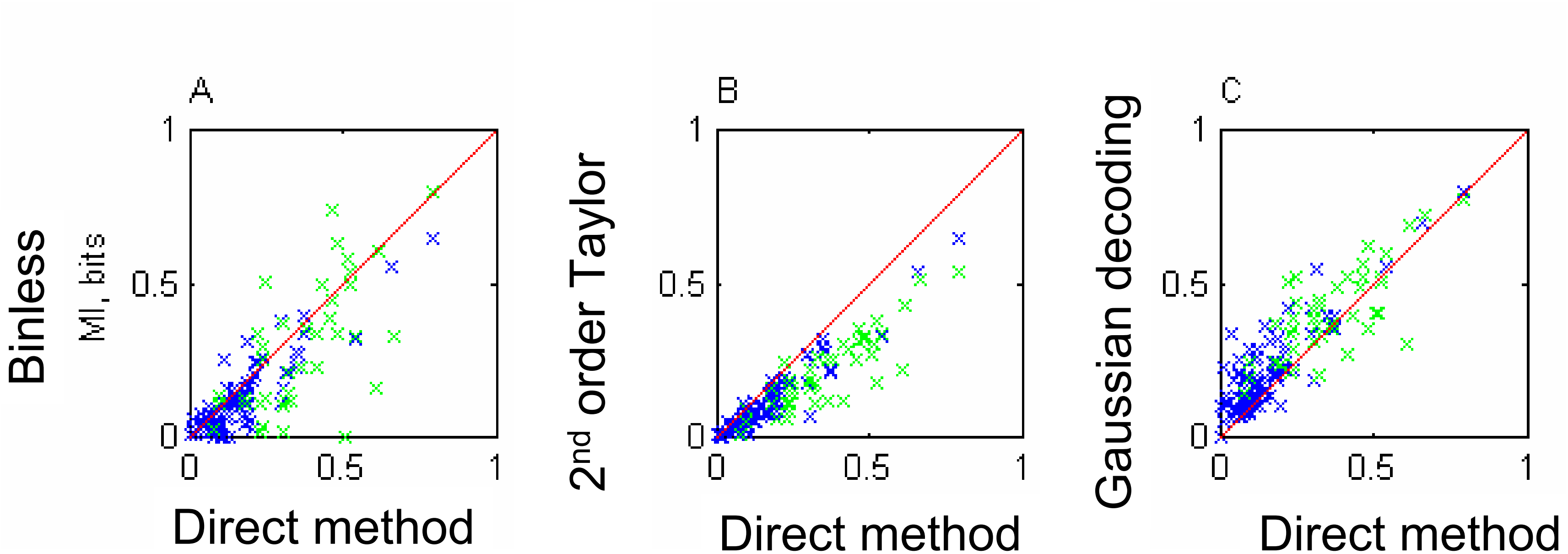


Outline

- General concerns in MI estimation
- Data and methods compared
- Validation of methods on simulated data
- **Results with real data**
- Simple statistics of spike trains
- Redundancy in the auditory pathway

In real data

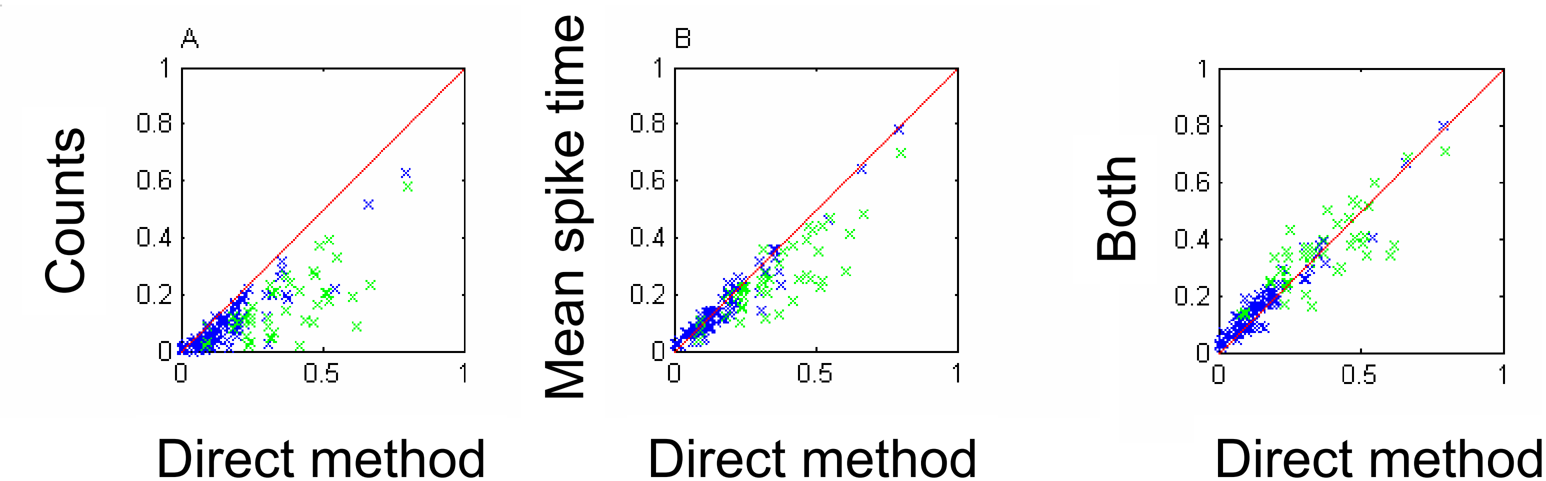
Estimating MI with the real data



- Green = ferrets, Blue = cats

Using simple statistics (real data)

- Could the same level of information be obtained with simple statistics?



- Spike counts and mean response time capture all information obtained with the direct estimator

Outline

- General concerns in MI estimation
- Data and methods compared
- Validation of methods on simulated data
- Results with real data
- Simple statistics of spike trains
- **Redundancy in the auditory pathway**

Recordings from three auditory processing stations

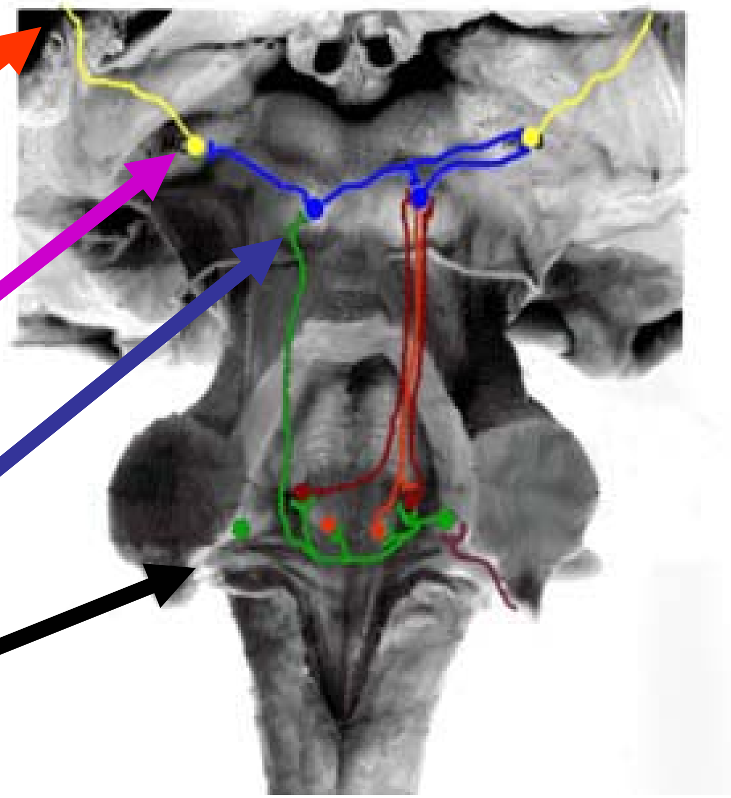
We record in Halothane anesthetized cats

Auditory cortex (AI)

Auditory Thalamus (MGB)

Auditory Mid Brain (IC)

Simulated ANF cells



Measure redundancies

- Pairwise redundancy:

$$I(R_1,R_2;S) - [\ I(R_1;S) + I(R_2;S)]$$

- Under conditional independence given the stimuli this equals: $= - I(R_1;R_2)$

- Redundancy tends to be correlated with single-cell MI $I(R;S)$, so we use normalized redundancies

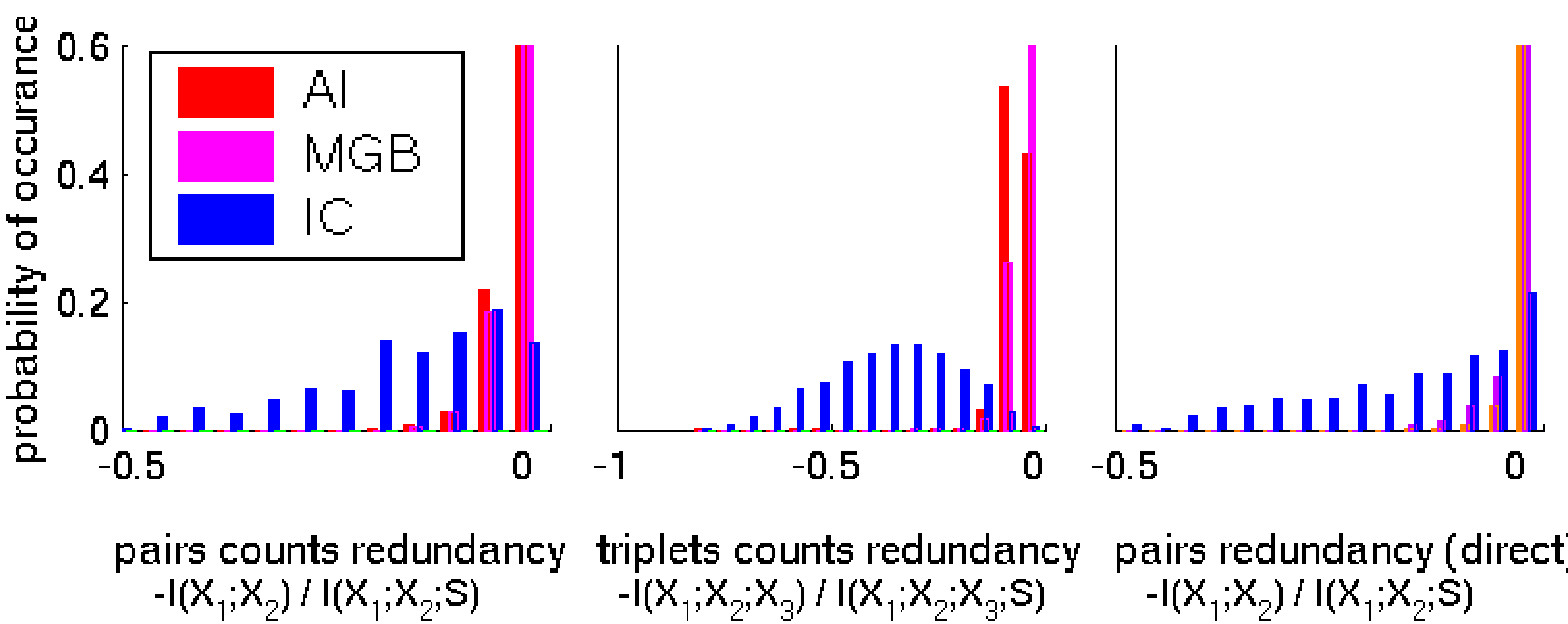
$$- I(R_1;R_2)\ /\ [\ I(R_1;S) + I(R_2;S)]$$

- Can be generalized to larger groups

Redundancy in the ascending pathway

Redundancy is reduced along the processing hierarchy:
Higher redundancy in IC but lower in MGB and AI, coding the same set of stimuli

- This is observed when estimating MI with
  - spike counts, first spike latency, the direct method.
- The effect is enhanced when considering triplets



Summary

- We compared four approaches for MI estimation:
  - Binless estimation [Victor]
  - 2nd order Taylor [Panzeri & Treves]
  - Direct method [Strong et al.]
  - Gaussian decoding [Treves]
- We evaluated their performance on simulated data that mimics the statistics of real recordings
- In simulated data, the direct method was the most accurate
- All information is the real data essentially captured by two simple statistics of the spike trains: spike counts and mean response time
- Reliable MI estimation reveals reduction in coding redundancy in the auditory processing hierarchy