Efficient Coding of Natural Sounds

Michael Lewicki

Center for the Neural Basis of Cognition &
Department of Computer Science

Carnegie Mellon University

How does the brain encode complex sensory signals?

Outline

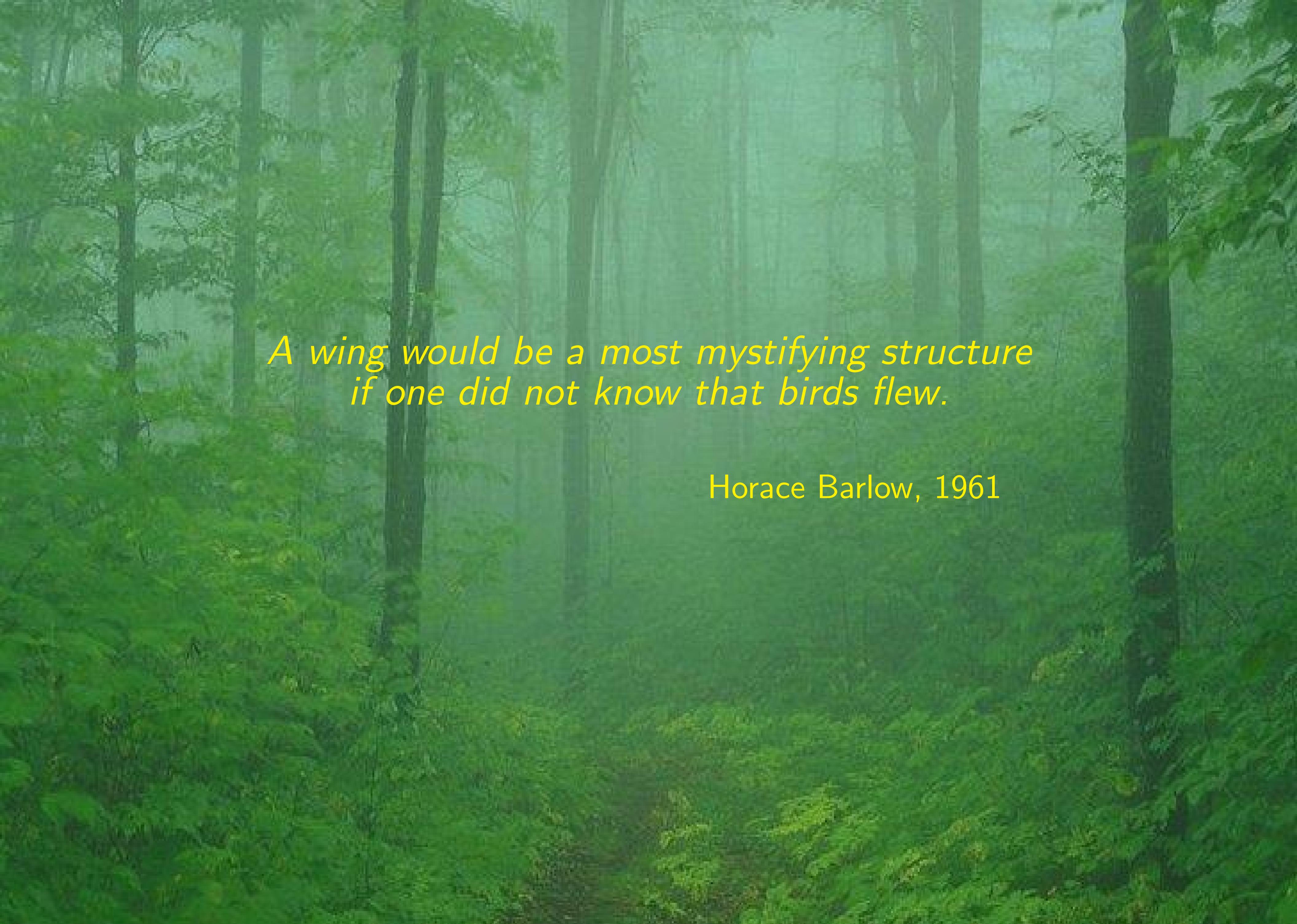
Motivations

Efficient coding theory

Application to natural sounds

Interpretation of experimental data

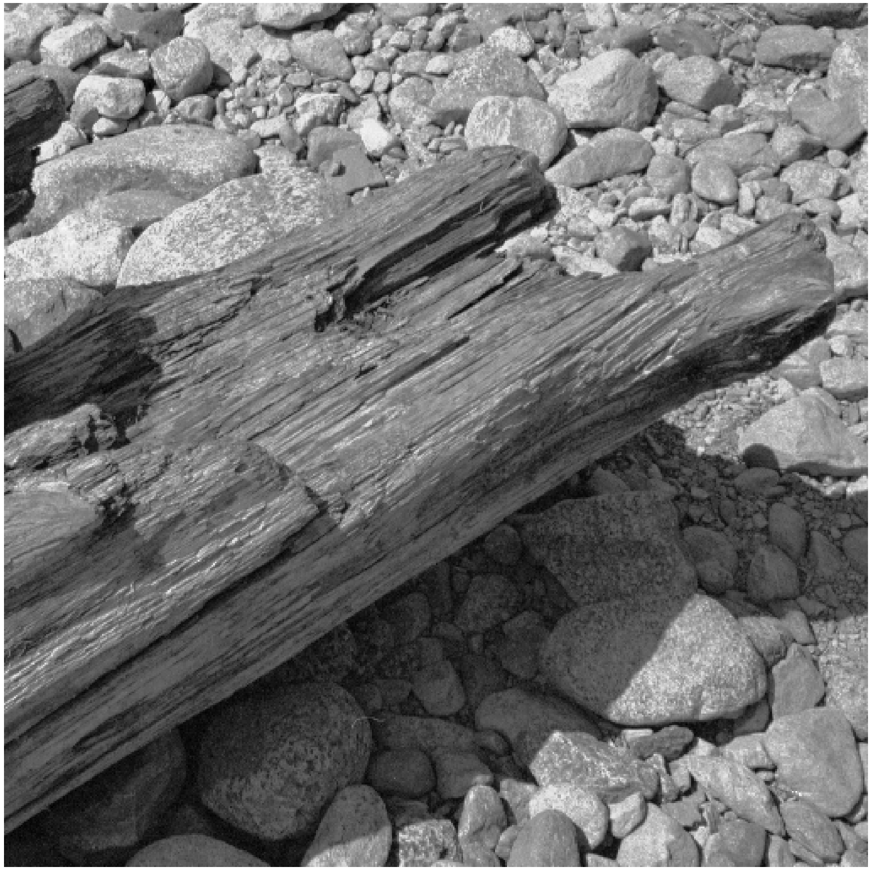
Efficient coding in population spike codes

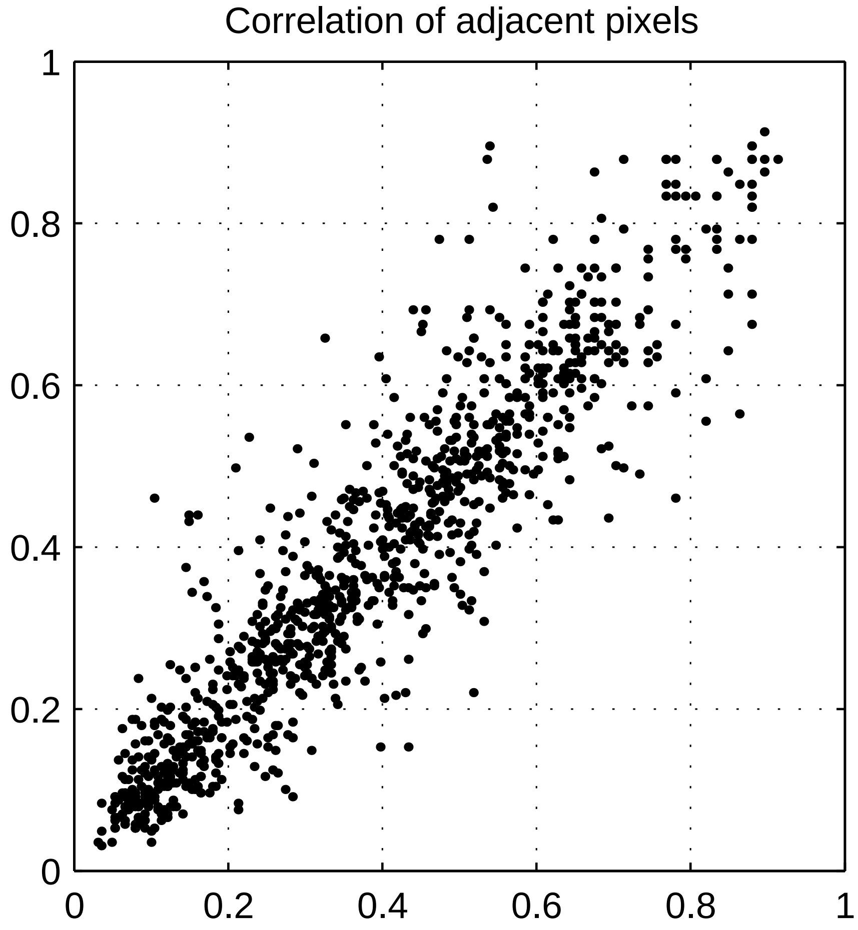


A wing would be a most mystifying structure if one did not know that birds flew.

Horace Barlow, 1961

Natural signals are redundant





Efficient coding hypothesis (Attneave, 1954; Barlow, 1961; et al):

Sensory systems encode only non-redundant structure

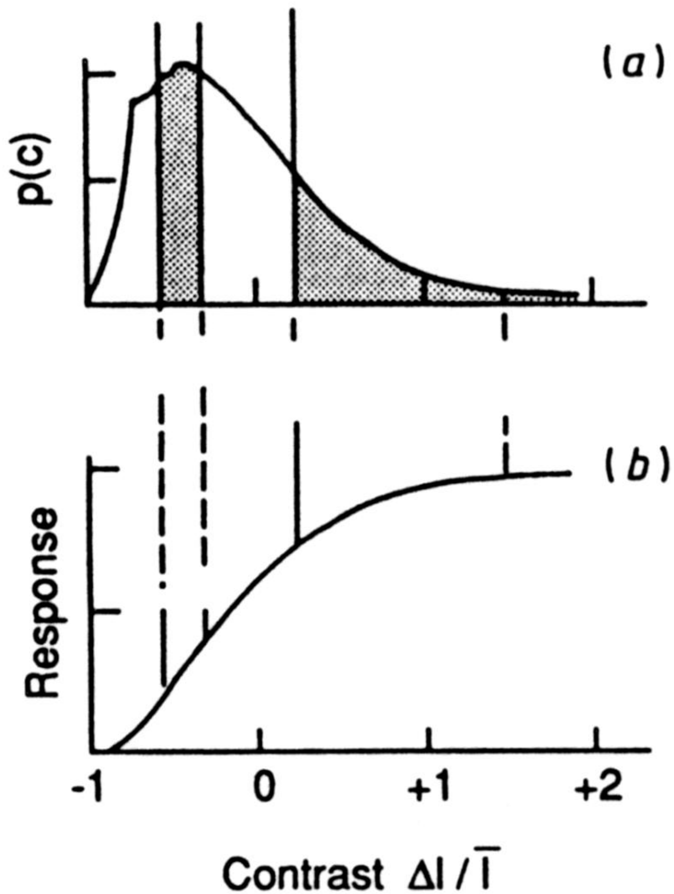
Why code efficiently?

Information bottleneck of sensory coding:

- restrictions on information flow rate
  - channel capacity of sensory nerves
  - computational bottleneck
  - $5 \times 10^6 \rightarrow 40 - 50$ bits/sec
- facilitate pattern recognition
  - independent features are more informative
  - better sensory codes could simply further processing
- other ideas
  - efficient energy use
  - faster processing time

How do we use this hypothesis to predict sensory codes?

A simple example: efficient coding of a single input



(from Atick, 1992)

How to set sensitivity?

- too high ⇒ response saturated
- too low ⇒ range under utilized

- inputs follow distribution of sensory environment
- encode so that output levels are used with equal frequency
- each response state has equal area (⇒ equal probability)
- continuum limit is cumulative pdf of input distribution

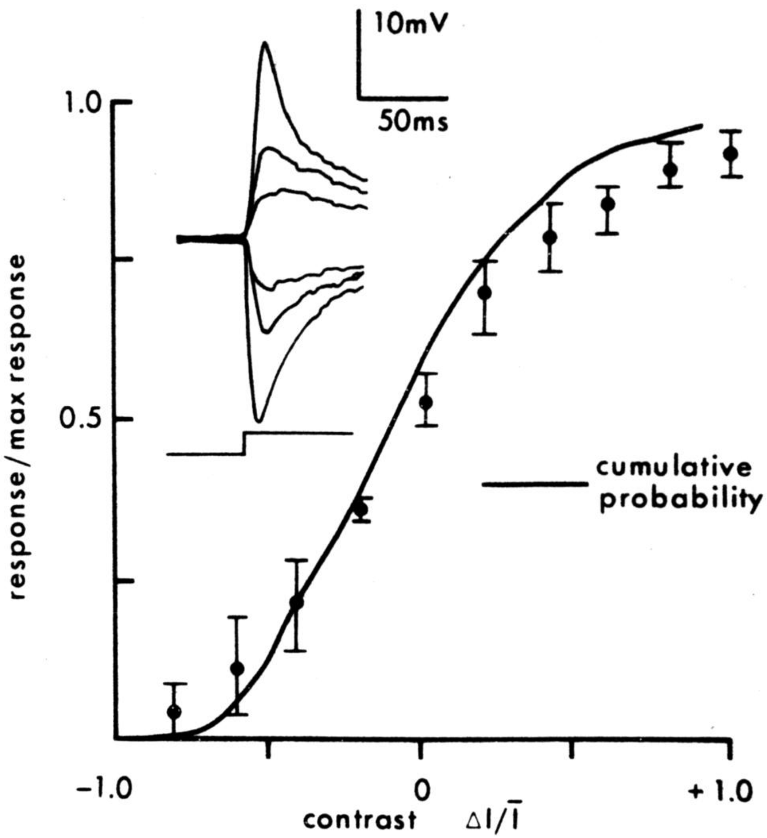
For $y = g(c)$

$$\frac{y}{y_{max}} = \int_{c_{min}}^{c} P(c')dc'$$

Testing the theory: Laugin, 1981

Laughlin, 1981:

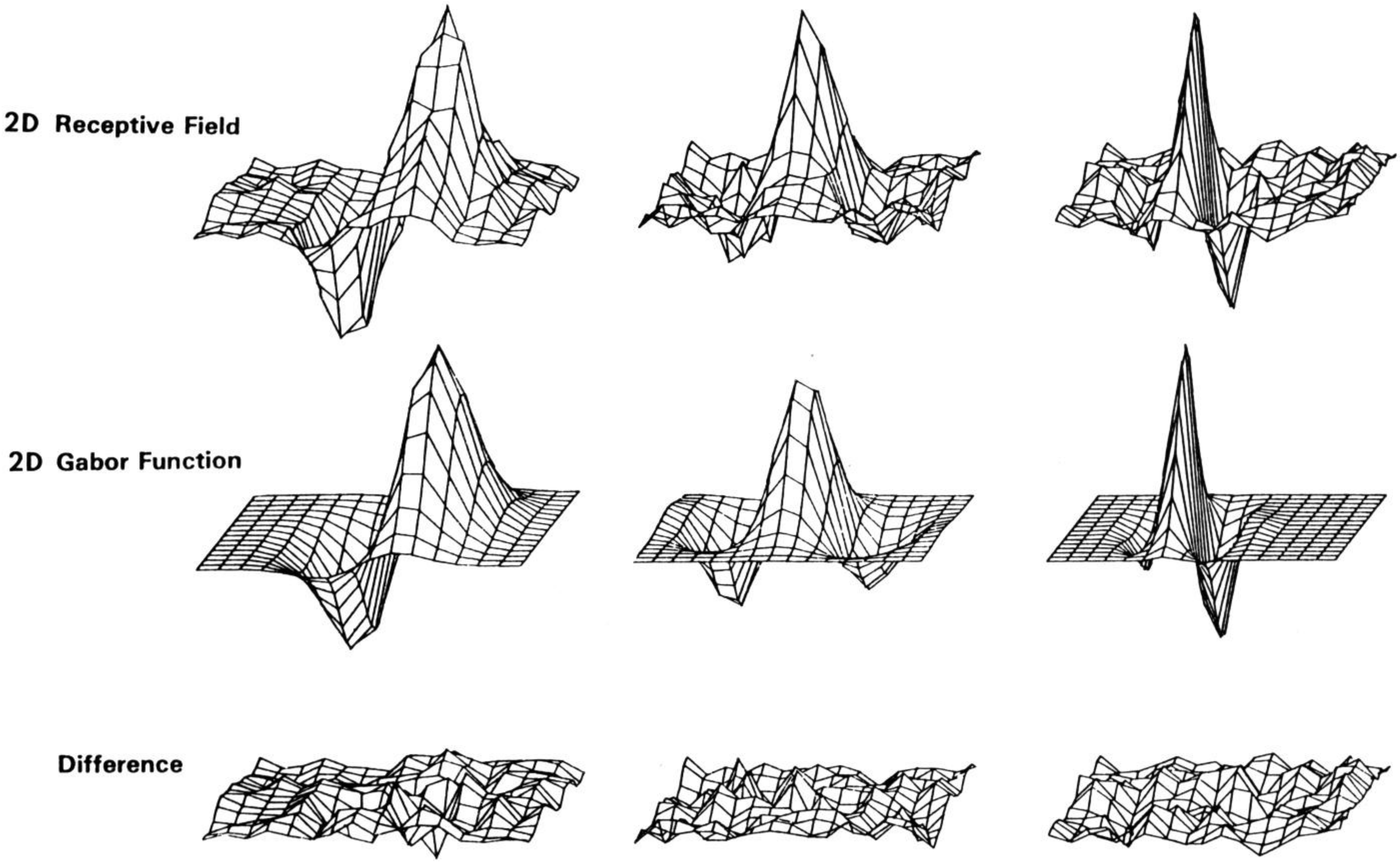
- predict response of fly LMC (large monopolar cells)
  – interneuron in compound eye
- output is graded potential



- collect natural scenes to estimate stimulus pdf
- predict contrast response function
  ⇒ fly LMC transmits information efficiently

What about complex sensory patterns?

V1 receptive fields are consistent with effcient coding theory

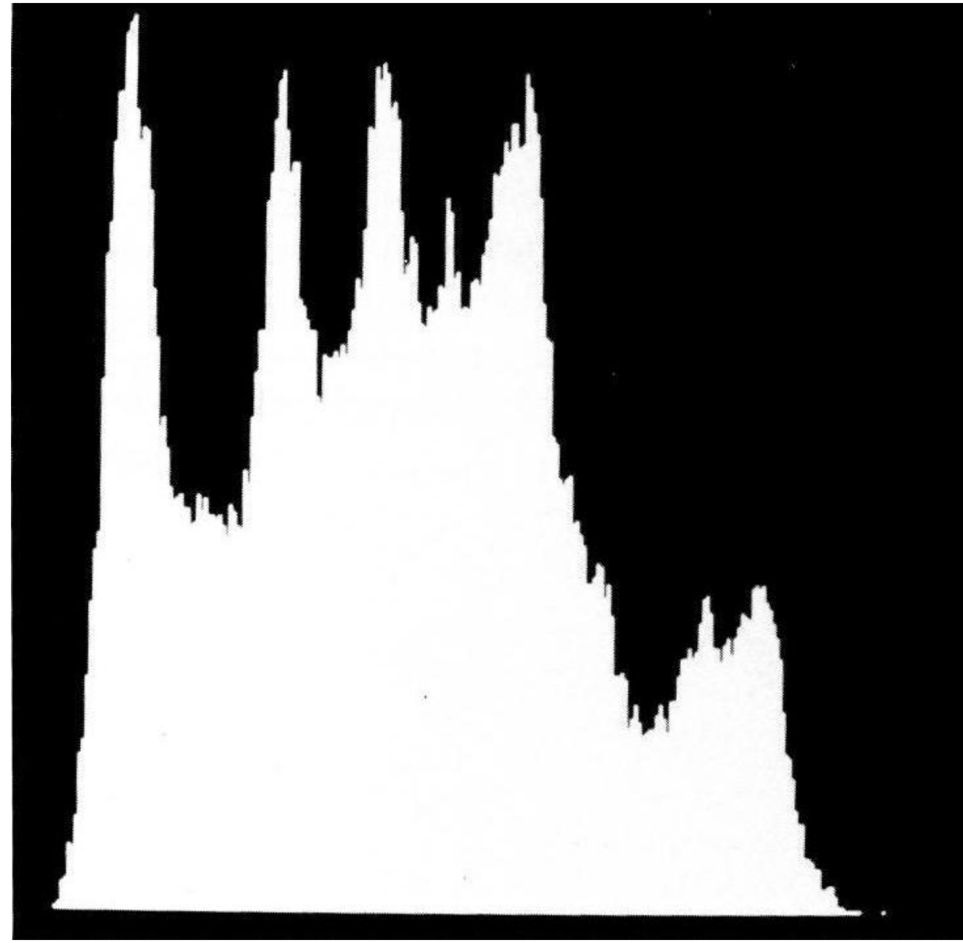


V1 receptive fields are well-fit by 2D Gabor functions (Jones and Palmer, 1987).

Does this yield an efficient code?

Coding images with pixels (Daugman, 1988)

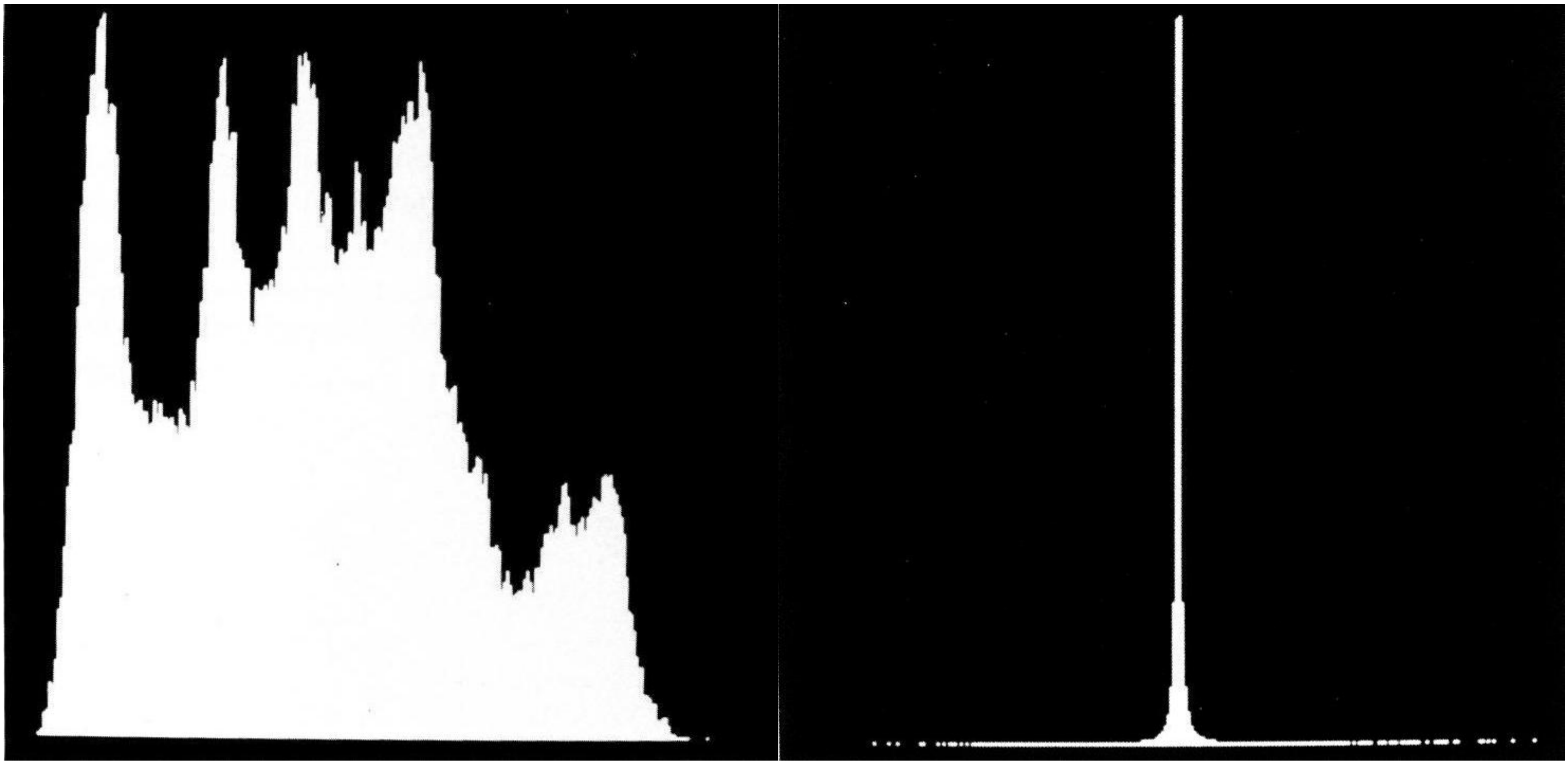
Lena



histogram of pixel values
Entropy = 7.57

High entropy means high redundacny ⇒ a very *inefficient* code

Recoding with Gabor functions (Daugman, 1988)

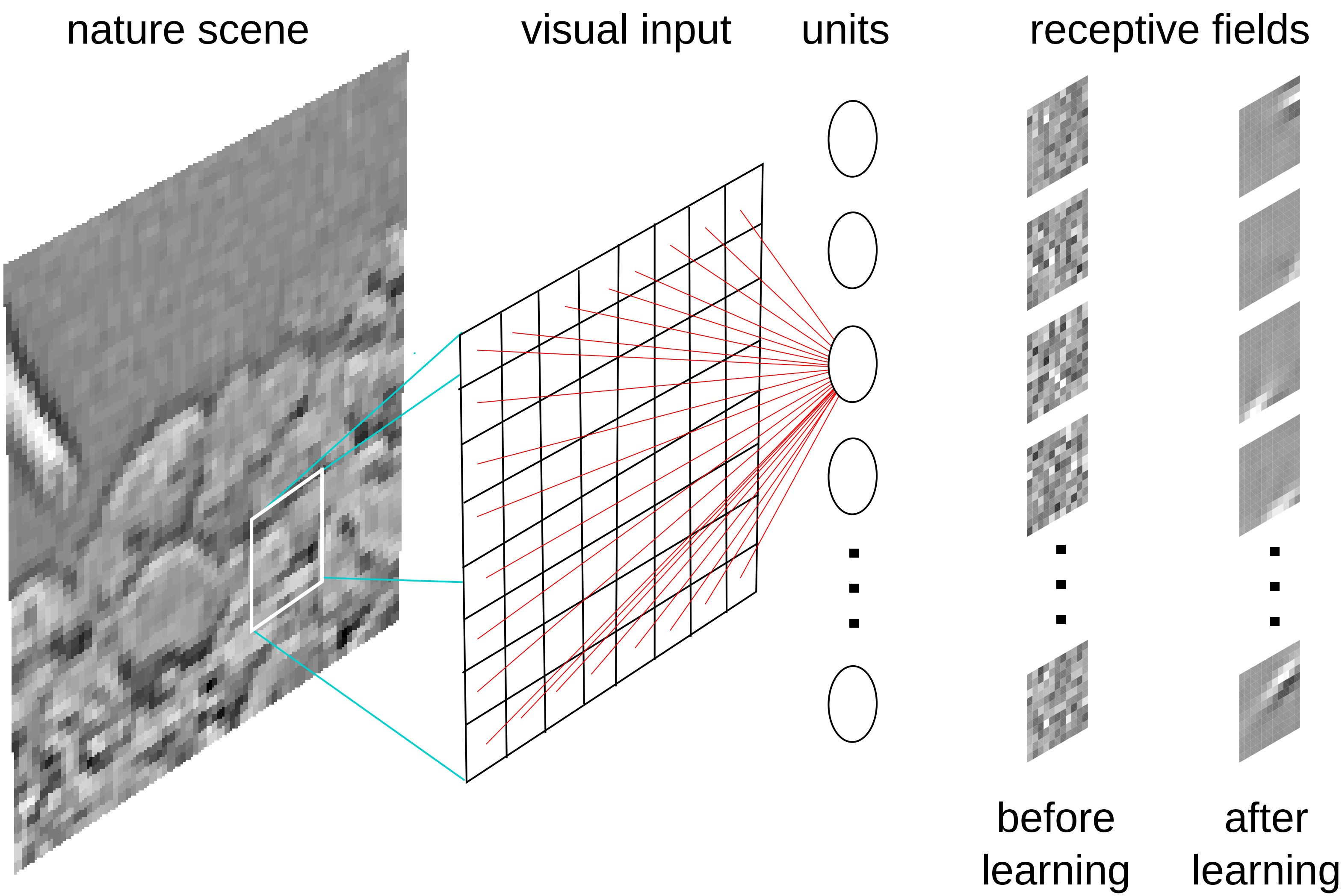


Pixel entropy= 7.57 bits

Recoding with 2D Gabor functions
Filter output entropy = 2.55 bits.

Can these codes be predicted?

Sparse coding of natural images (Olshausen and Field, 1996)

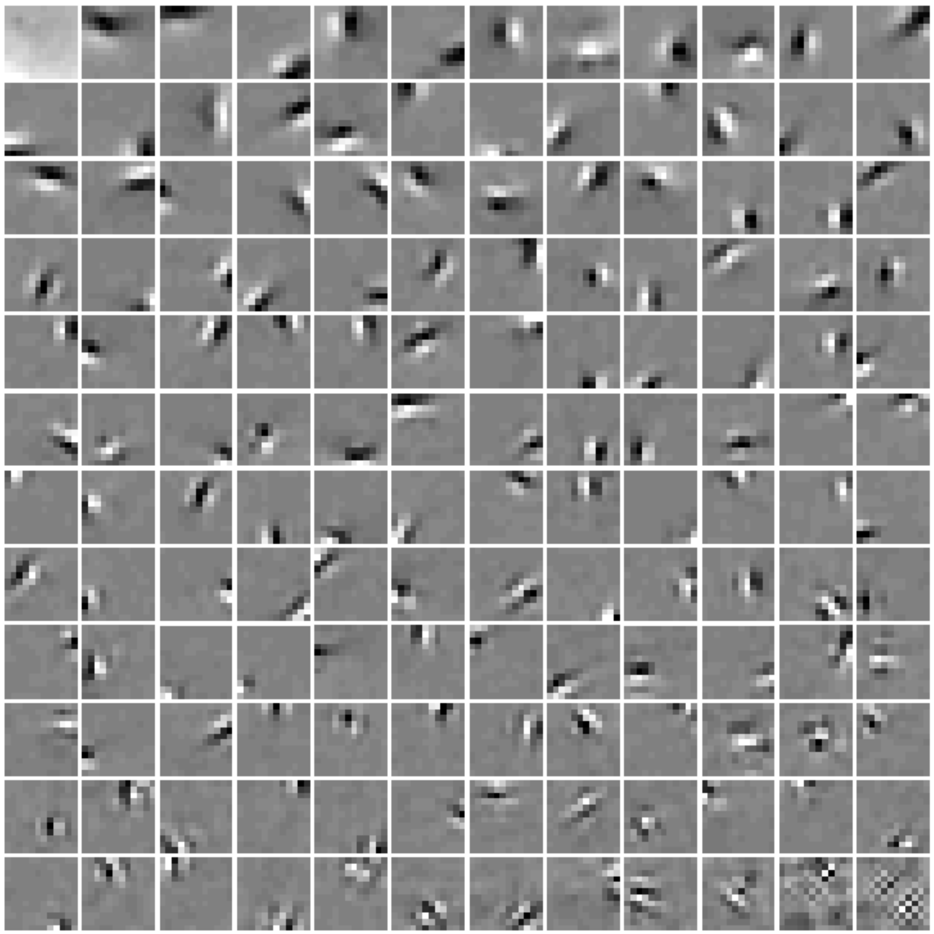


Adapt population of receptive fields to

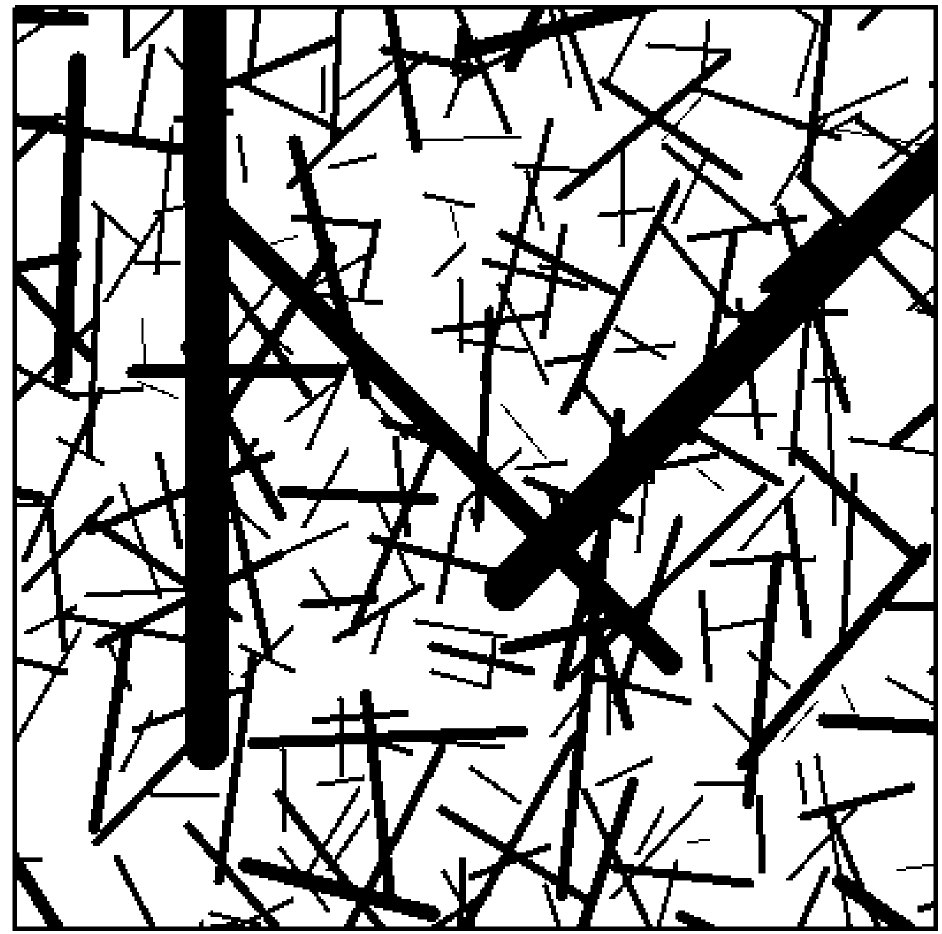
- accurately encode an ensembe of natural images
- maximizing the sparseness of the output, i.e. minimizing entropy.

Theory predicts entire population of receptive fields

(Lewicki and Olshausen, 1999)

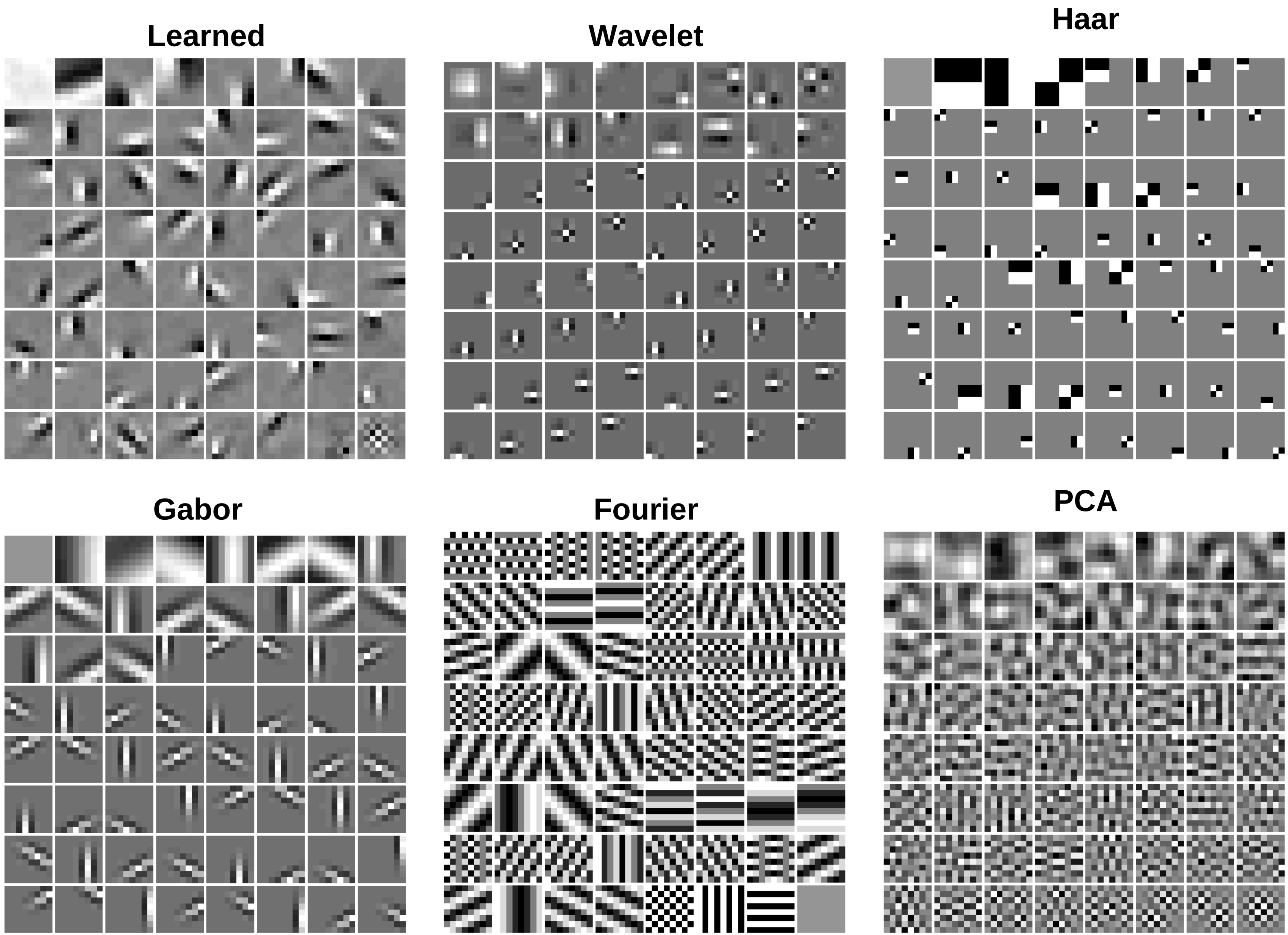


Population of receptive fields.
(black = inhibitory; white = excitatory)

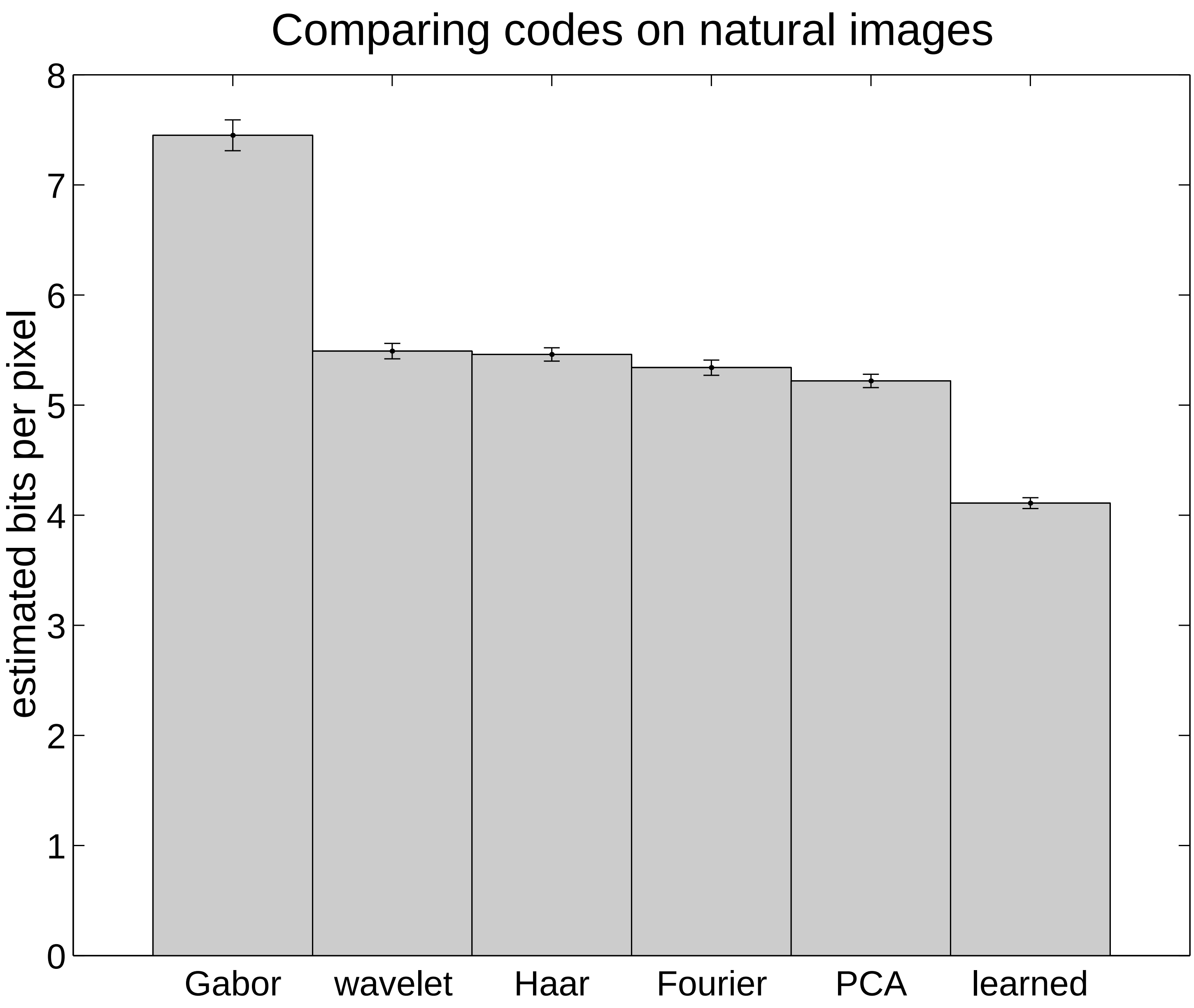


Overlayed response property schematics.

Algorithm selects best of many possible sensory codes

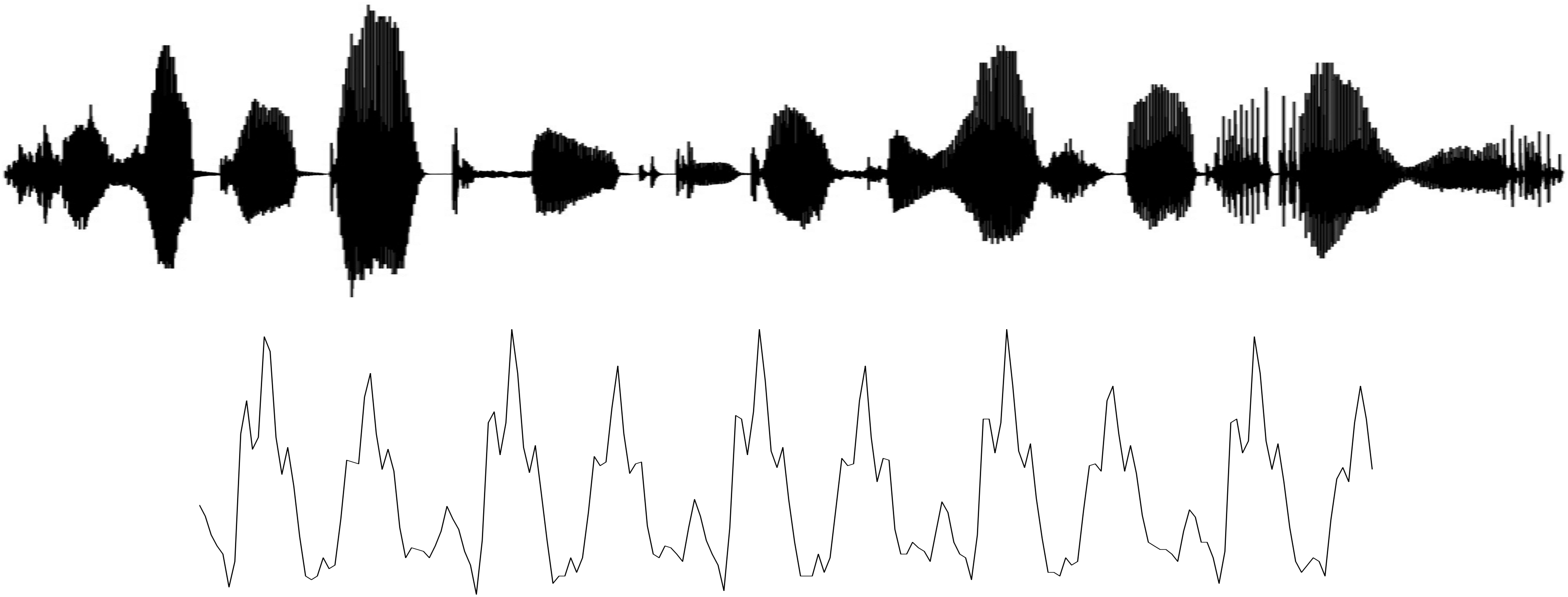


(Lewicki and Olshausen, 1999) Theoretical perspective:
Not edge "detectors" but an efficient way to describe natural, complex images.


Comparing codes on natural images
estimated bits per pixel
0
1
2
3
4
5
6
7
8
Gabor
wavelet
Haar
Fourier
PCA
learned

Efficient coding of natural sounds

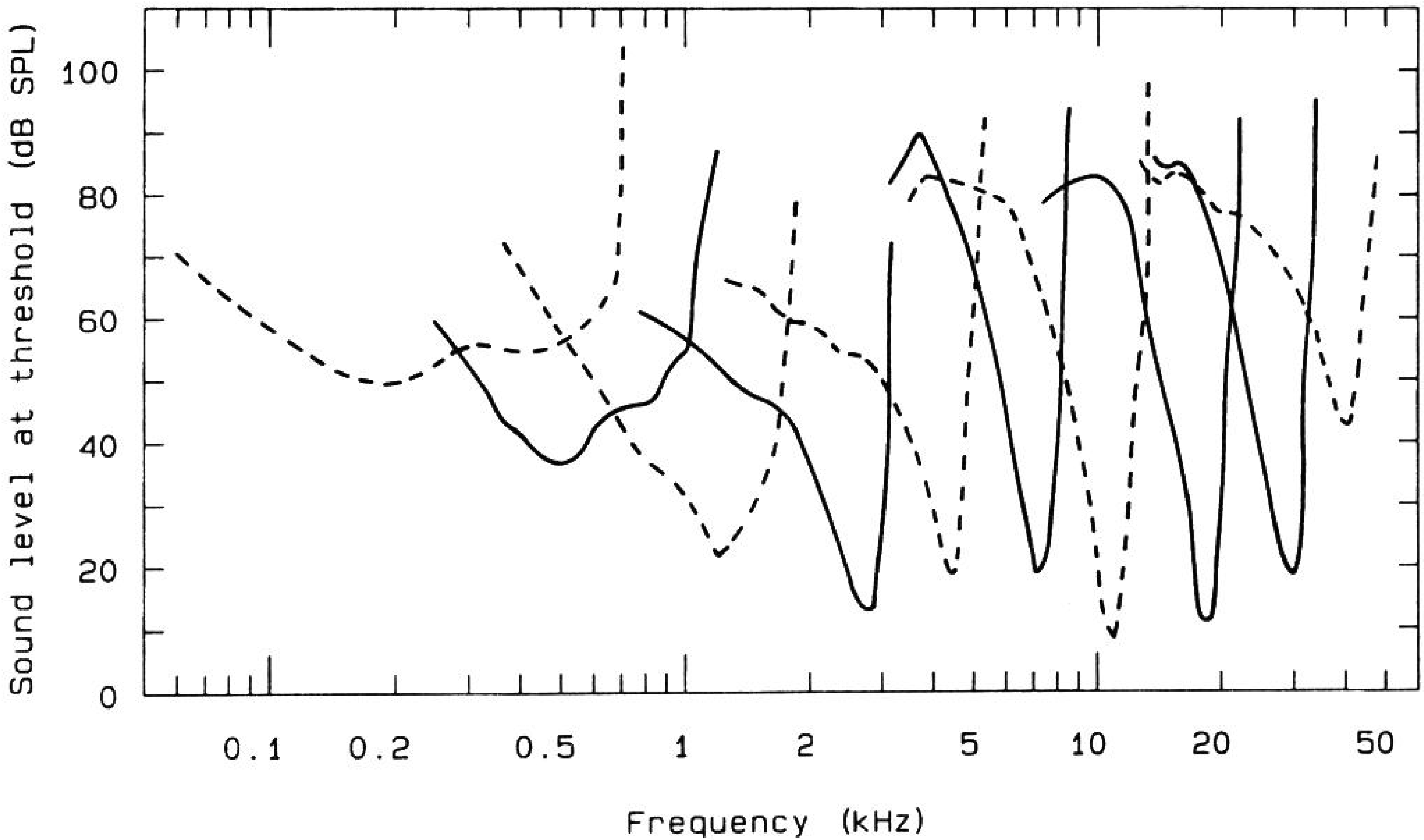
Efficient coding: focus on coding waveform directly



Goal:

Predict optimal transformation of acoutsic waveform from statistics of the acoustic environment.

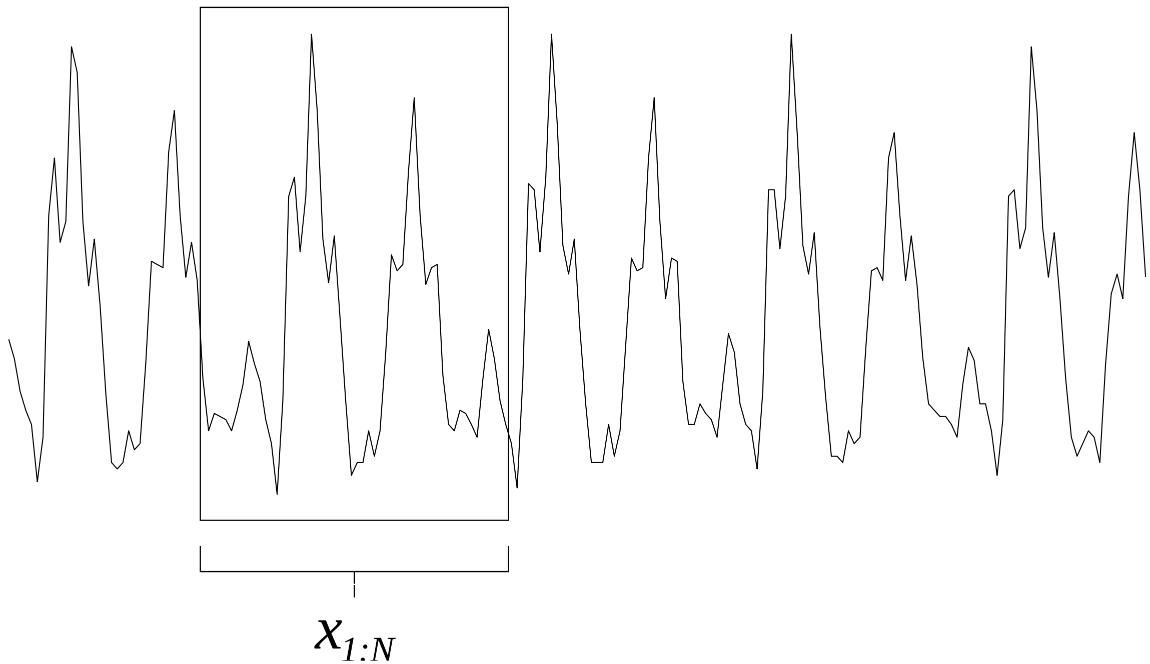
Why encode sound by frequency?



Auditory tuning curves.

A simple model of waveform encoding

Data consists of waveform segments sampled randomly from a sound ensemble:



Filterbank model:

$$a_i(t) = \sum_{\tau=0}^{N-1} x(t-\tau) h_i(\tau)$$

How do derive the filter shapes $h_i(t)$ that optimize coding efficiency?

Model only describes signals within the window of analysis.

Information theoretic viewpoint

Use Shannon's source coding theorm.

$$
\begin{aligned}
\mathcal{L} = E[l(X)] &\geq \sum_x p(x) \log \frac{1}{q(x)} \\
&= \sum_x p(x) \log \frac{p(x)}{q(x)} + \sum_x p(x) \log \frac{1}{p(x)} \\
&= D_{KL}(p \| q) + H(p)
\end{aligned}
$$

If model density $q(x)$ equals true density $p(x)$ then $D_{KL} = 0$.
$\Rightarrow q(x)$ gives *lower bound* on average code length.

greater coding efficiency $\Leftrightarrow$ *more learned structure*

Principle

Good codes capture the statistical distribution of sensory patterns.

How do we descibe the distribution?

Describing signals with a simple statistical model

Goal is to *encode* the data to desired precision

$$\begin{aligned}\mathbf{x} &= \vec{a}_1 s_1 + \vec{a}_2 s_2 + \cdots + \vec{a}_L s_L + \vec{\epsilon} \\ &= \mathbf{A}\mathbf{s} + \boldsymbol{\epsilon}\end{aligned}$$

Can solve for $\hat{\mathbf{s}}$ in the no noise case

$$\hat{\mathbf{s}} = \mathbf{A}^{-1}\mathbf{x}$$

Want algorithm to choose optimal $\mathbf{A}$ (i.e. the basis matrix).

Algorithm for deriving efficient codes

Learning objective:

maximize coding efficiency

$\Rightarrow$ maximize $P(\mathbf{x}|\mathbf{A})$ over $\mathbf{A}$ (basis for analysis window, or filter shapes).

Probability of pattern ensemble is:

$$P(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N|\mathbf{A}) = \prod_k P(\mathbf{x}_k|\mathbf{A})$$

To obtain $P(\mathbf{x}|\mathbf{A})$ marginalize over $\mathbf{s}$:

$$\begin{aligned} P(\mathbf{x}|\mathbf{A}) &= \int d\mathbf{s}\, P(\mathbf{x}|\mathbf{A}, \mathbf{s}) P(\mathbf{s}) \\ &= \frac{P(\mathbf{s})}{|\det \mathbf{A}|} \end{aligned}$$

Using *independent component analysis* (ICA) to optimize $\mathbf{A}$:

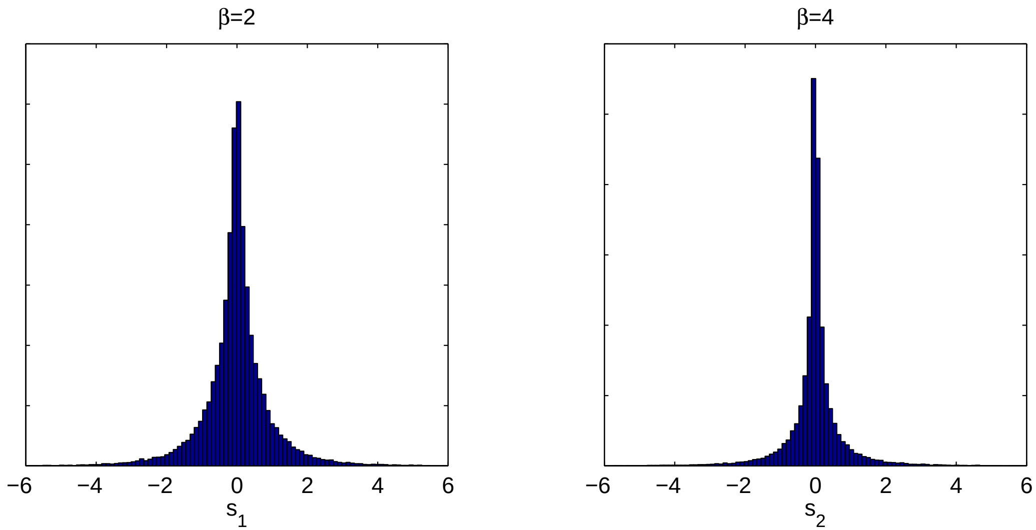
$$\begin{aligned} \Delta\mathbf{A} &\propto \mathbf{A}\mathbf{A}^T \frac{\partial}{\partial \mathbf{A}} \log P(\mathbf{x}|\mathbf{A}) \\ &= -\mathbf{A}(\mathbf{z}\mathbf{s}^T - \mathbf{I}), \end{aligned}$$

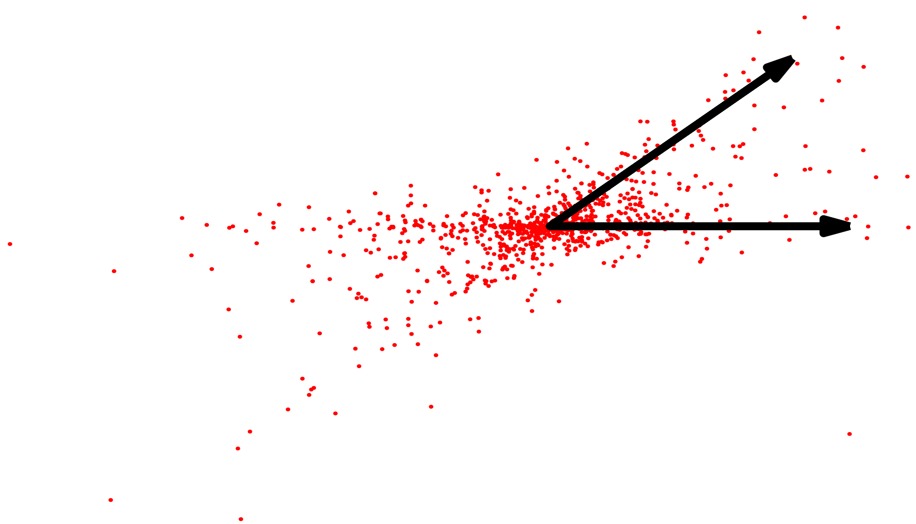
where $\mathbf{z} = (\log P(\mathbf{s}))'$. Use $P(s_i) \sim \text{ExPwr}(s_i|\mu, \sigma, \beta_i)$.

This learning rule:

- learns features that capture the most structure
- optimizes the efficiency of the code

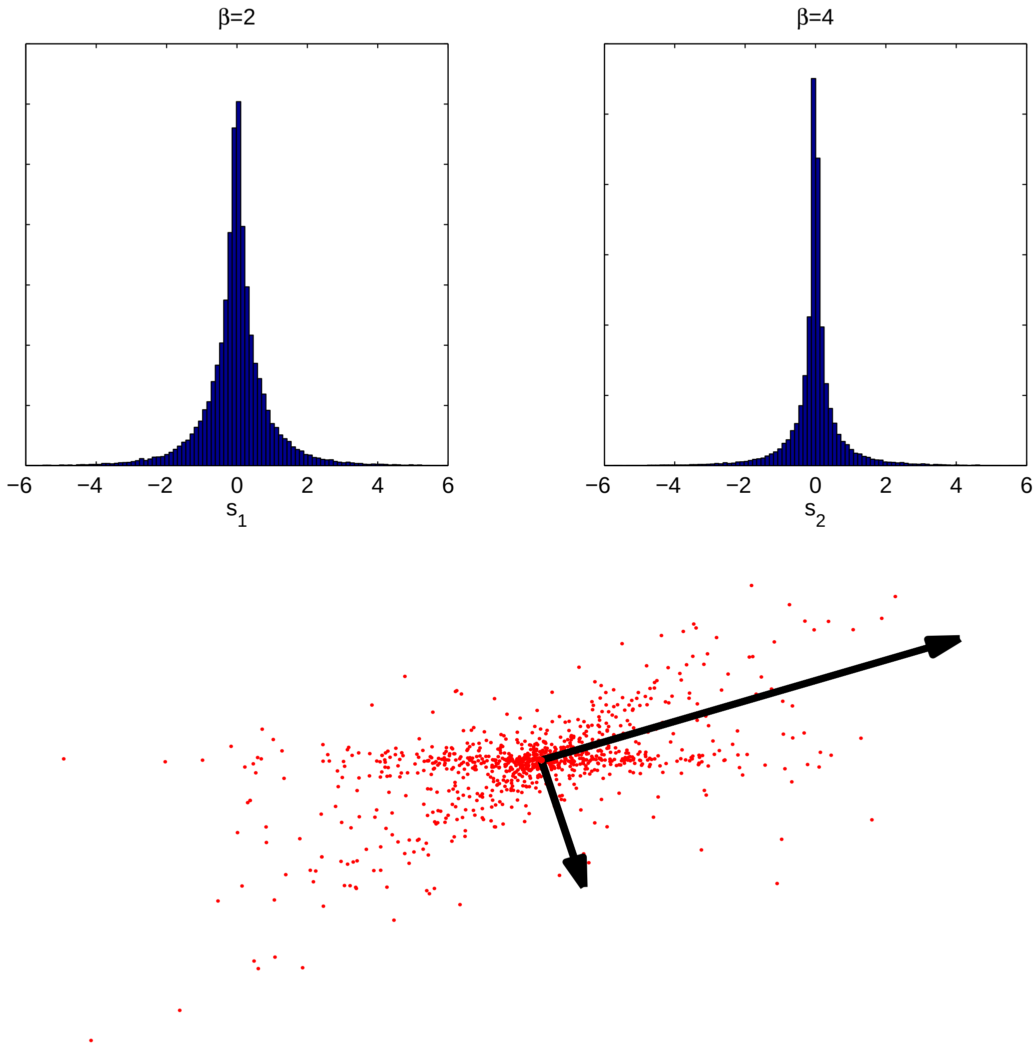
Modeling Non-Gaussian distributions with ICA





- Typical coeff. distributions of natural signals are *non-Gaussian*.
- Independent component analysis (ICA) describes the statistical distribution of non-Gaussian distributions
- The distribution is fit by optimizing the filter shapes.
- Unlike PCA, vectors are not restricted to be orthogonal.
- This permits a much better description of the actual distribution of natural signals.

Modeling Non-Gaussian distributions with ICA



- Typical coeff. distributions of natural signals are *non-Gaussian*.
- Independent component analysis (ICA) describes the statistical distribution of non-Gaussian distributions
- The distribution is fit by optimizing the filter shapes.
- Unlike PCA, vectors are not restricted to be orthogonal.
- This permits a much better description of the actual distribution of natural signals.

Efficient coding of natural sounds: Learning procedure

To derive the filters:

- select sound segments randomly from sound ensemble
- optimize filter shapes to maximize coding efficiency

What sounds should we use?

What are auditory systems adapted for?

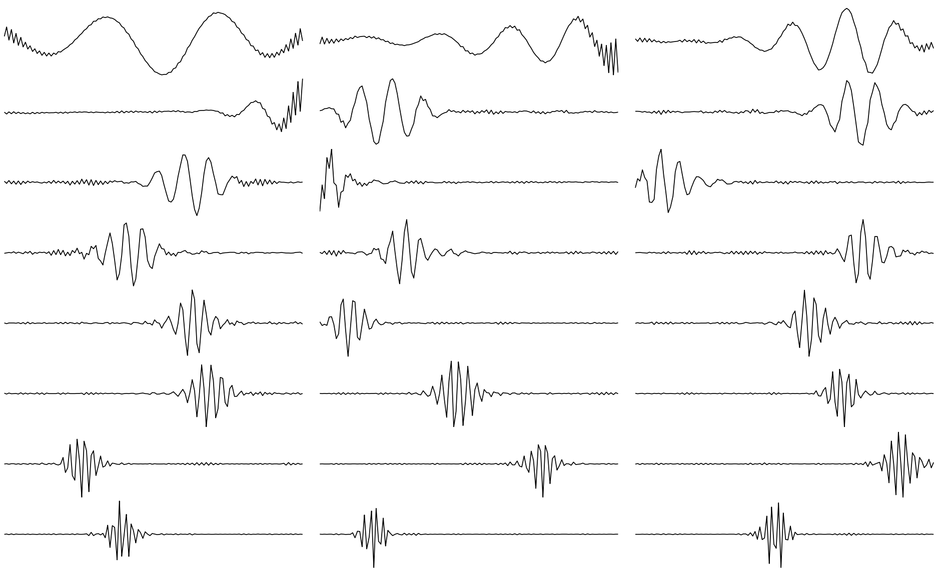
- localization / environmental sounds?
- communication / vocalizations?
- specific tasks, e.g sound discrimination?

We used the following sound ensembles:

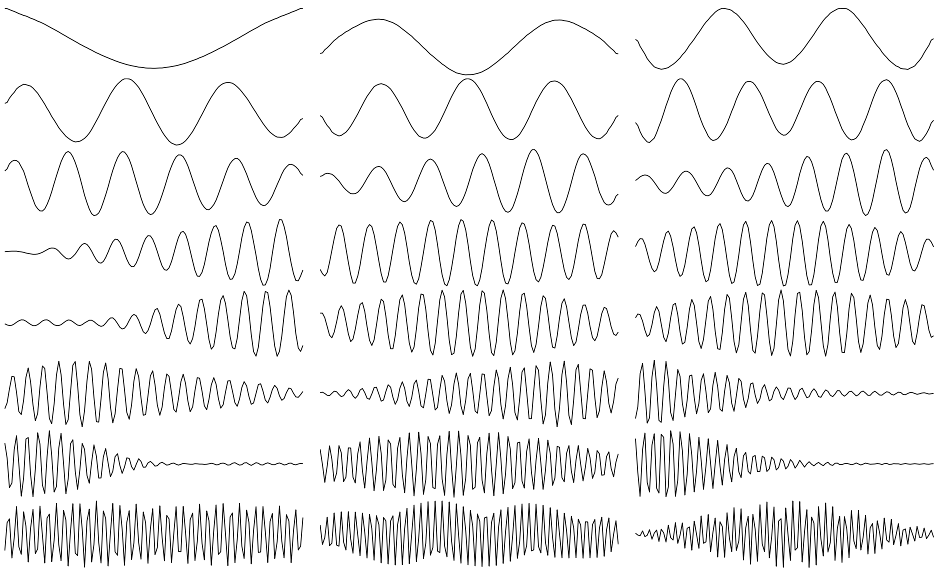
- non-harmonic environmental sounds (e.g. footsteps, stream sounds, etc.)
- animal vocalizations (rainforest mammals, e.g chirping, screeching, cries, etc.)
- speech (samples from 100 male and female speakers from the TIMIT corpus)

Results of adapting filters to different sound classes

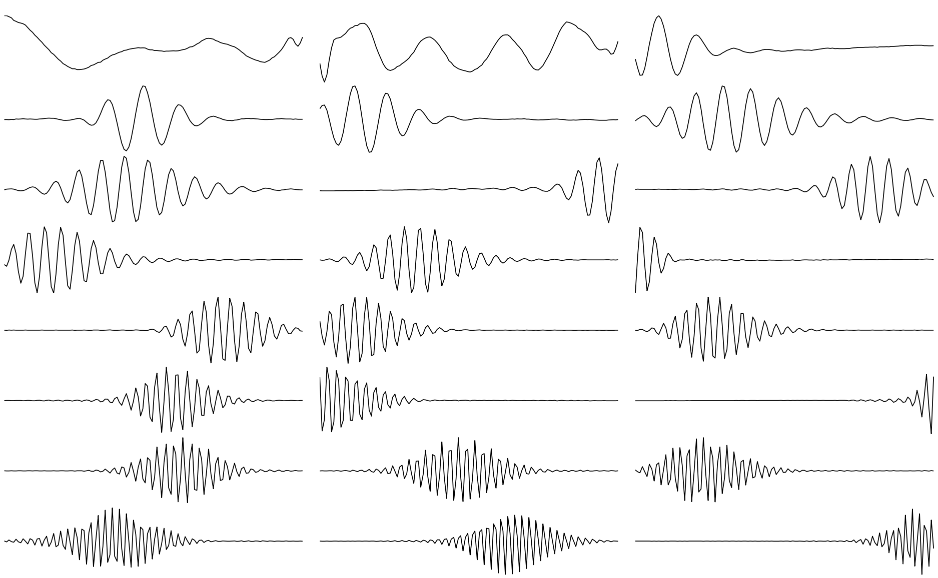
Efficient filters for environmental sounds:



Efficient filters for animal vocalizations:

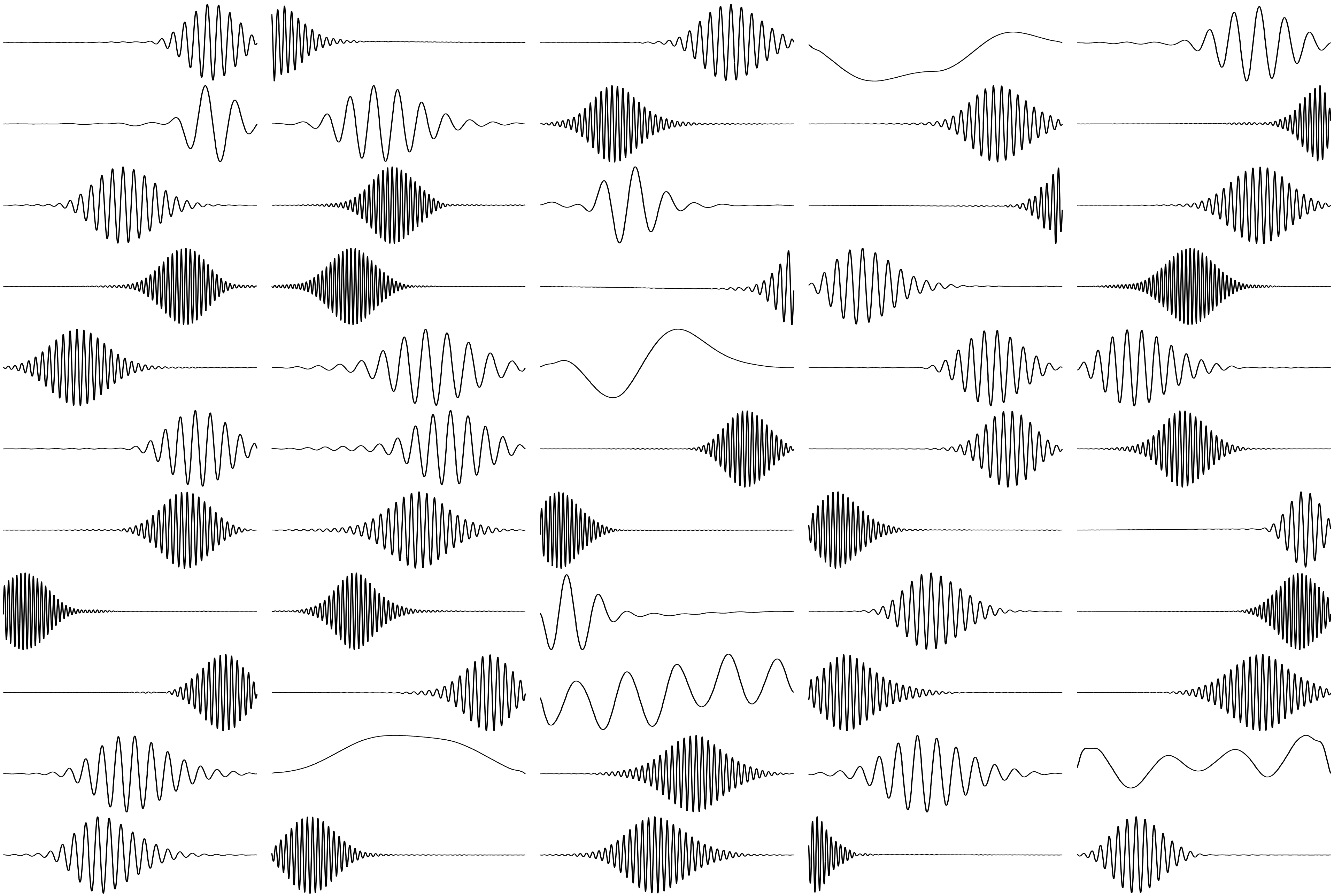


Efficient filters for speech:



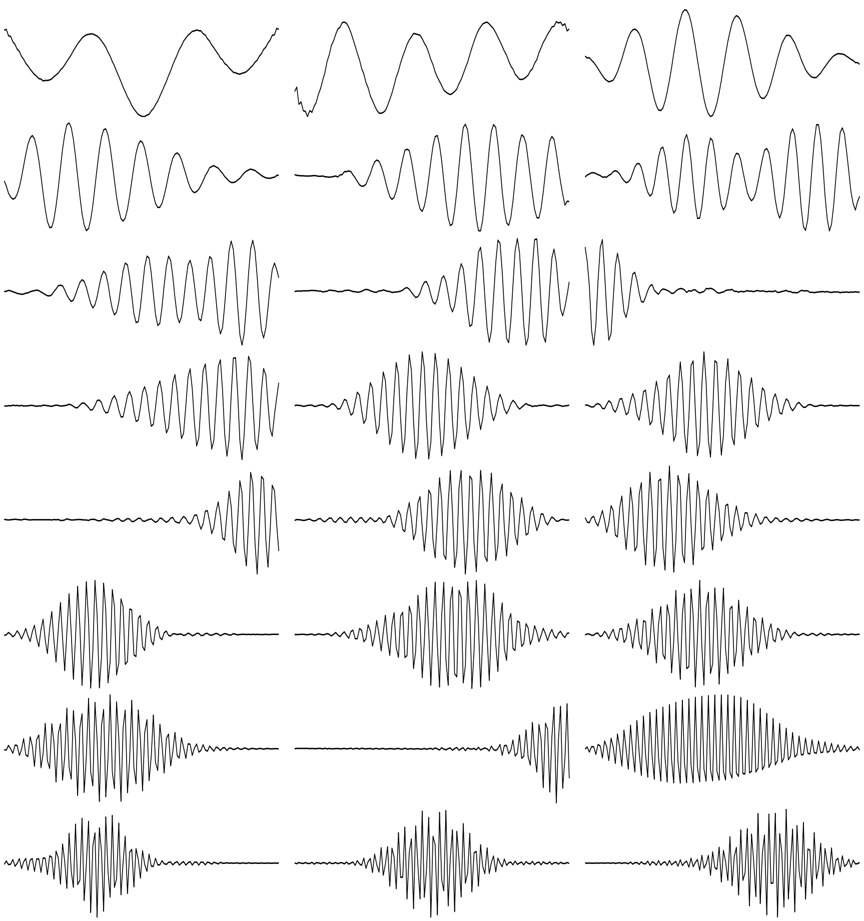
- Each result shows only a subset
- Auditory nerve filters best match those derived from environmental sounds and speech
- learning movie

Upsampling removings aliasing due to periodic sampling

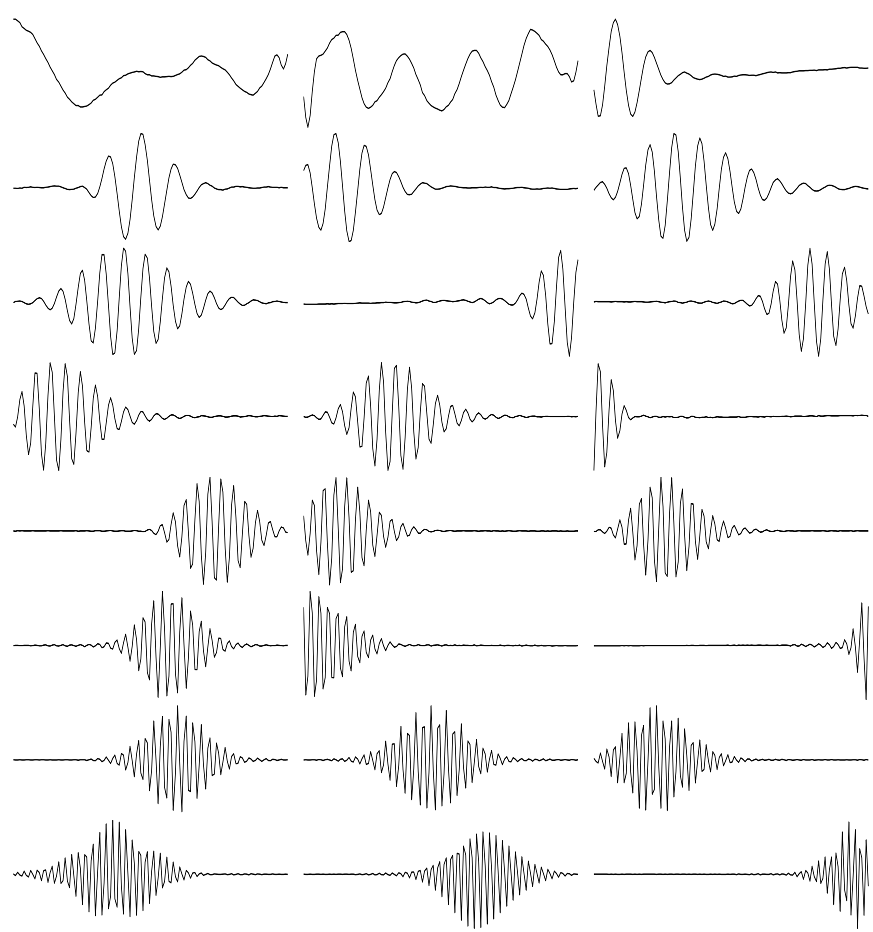


A combined ensemble: env. sounds and vocalizations

Efficient filters for combined



Efficient filters for speech:



Can vary along the continuum by changing relative proportion, best match is 2:1
⇒ speech is well-matched to the auditory code

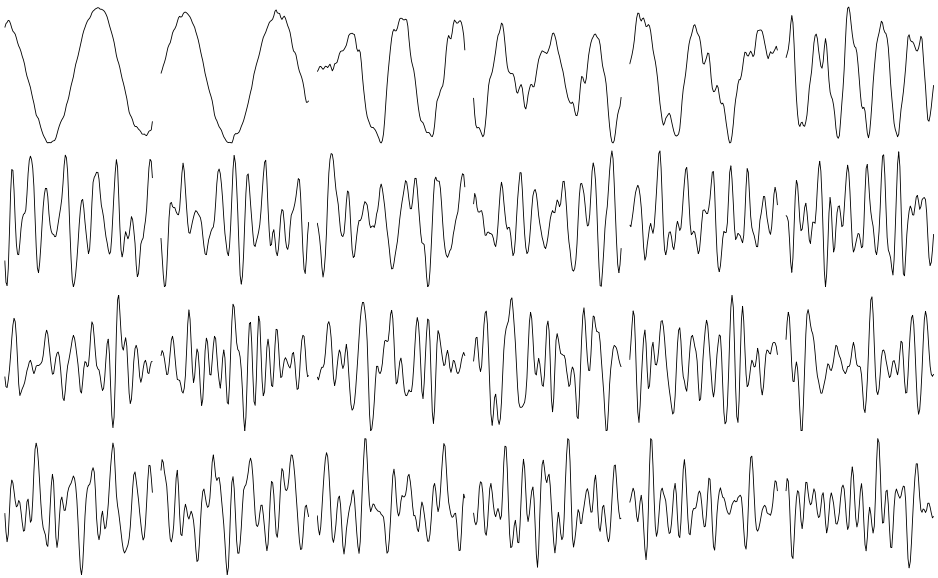
Can decorrelating models also explain data?

Redundancy reduction models that adapt weights to decorrelate output activies assume a Gaussian model:

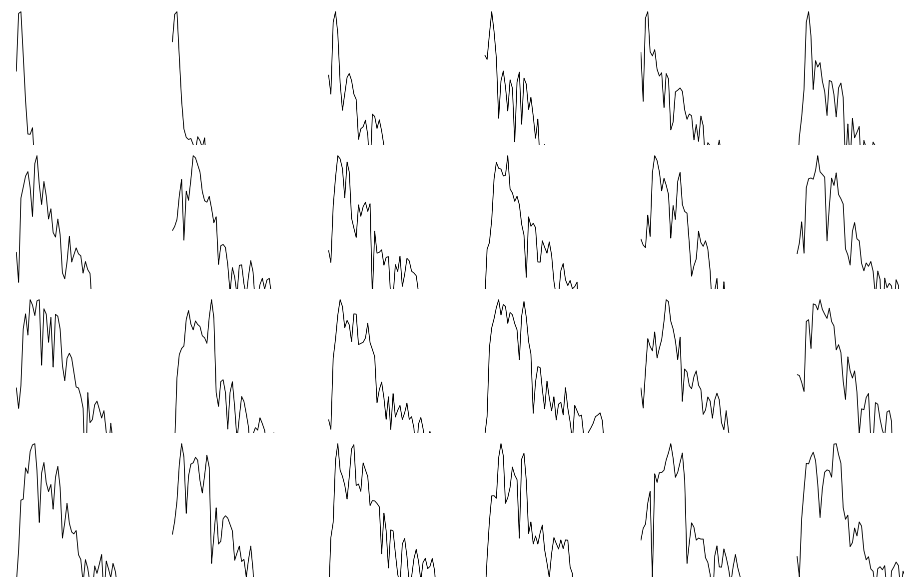
$$\mathbf{x} \sim \mathcal{N}(\mathbf{x}|\mu, \sigma)$$

Under this model, the filters can be derived with principal component analysis.

PCs of Environmental Sounds:



Corresponding Power Spectra:



$\Rightarrow$ just decorrelating the outputs does not yield time-frequency localized filters.

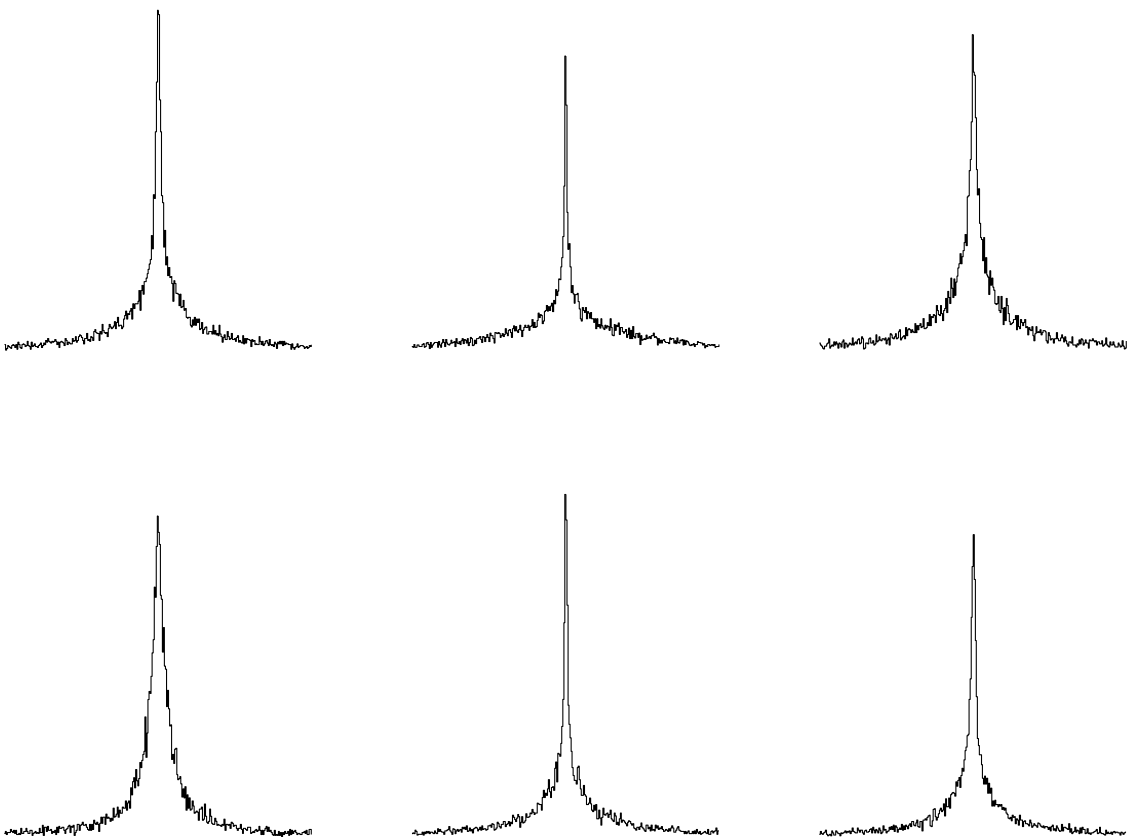
Why doesn't PCA work?

Check assumptions:

$\mathbf{x} = \mathbf{A}\mathbf{s}$ and $\mathbf{x} \sim \mathcal{N}(\mathbf{x}|\mu, \sigma)$

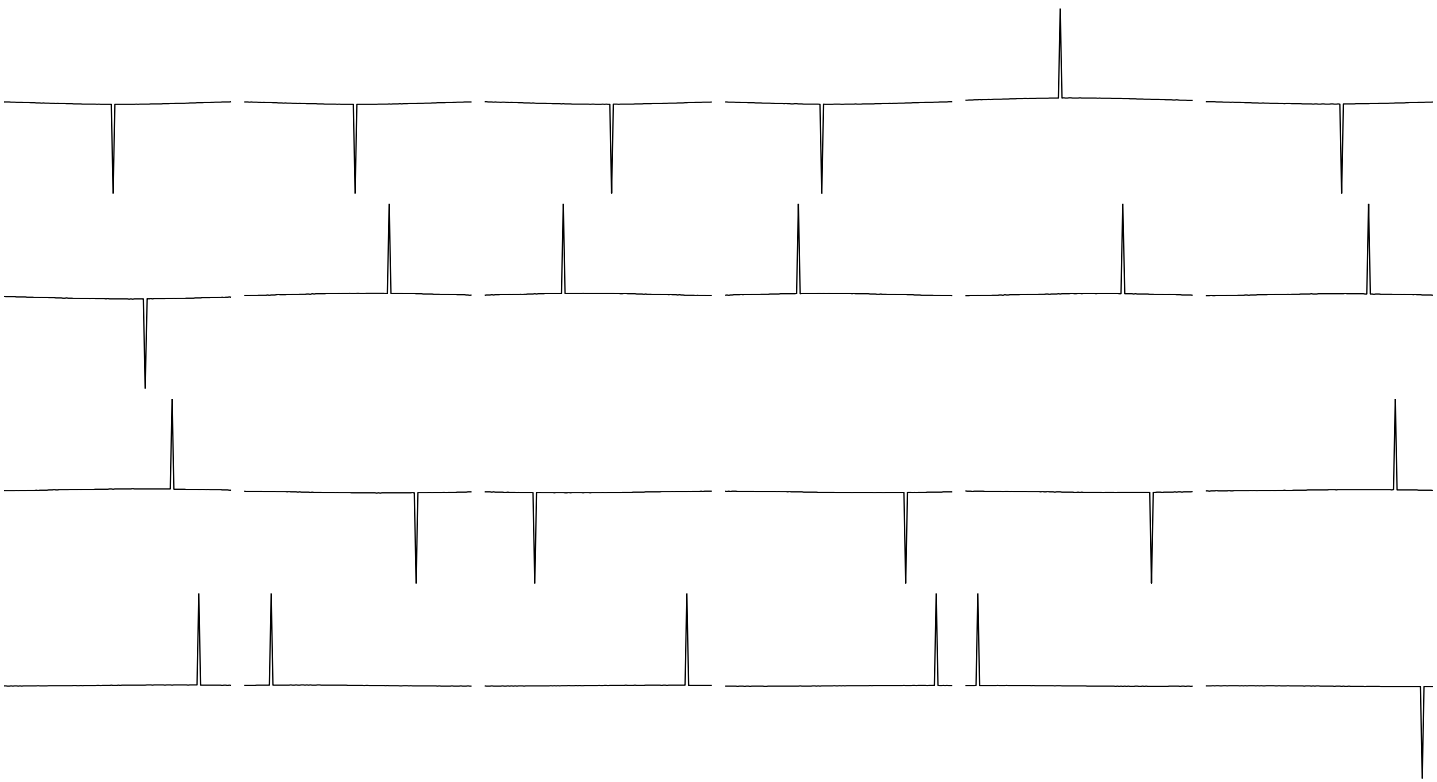
$\Rightarrow$ distribution of $\mathbf{s}$ should also be Gaussian.

Actual distribution of filter coefficients:



Efficient coding of sparse noise

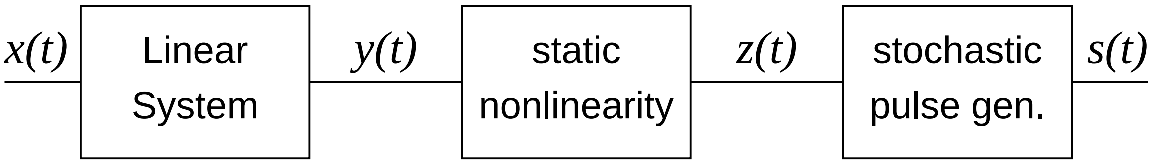
Learned sparse noise filters:



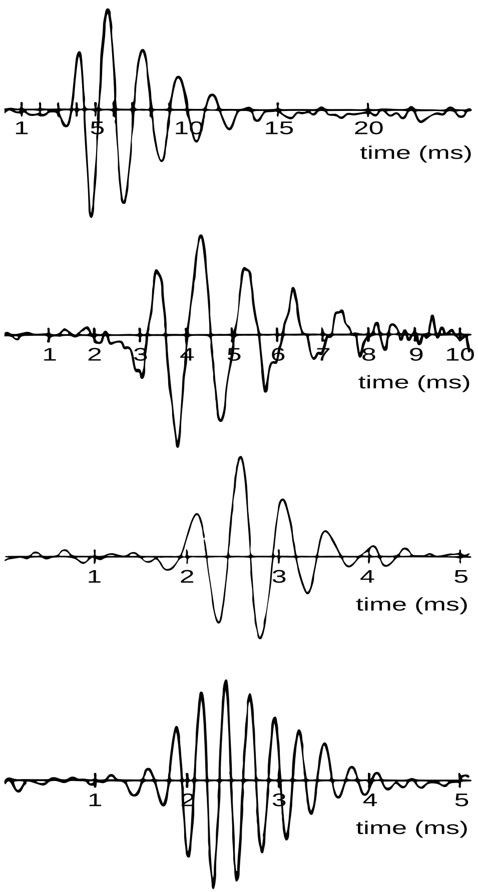
Efficient filters are delta functions that represent different time points in the analysis window.

...but what about the auditory system?

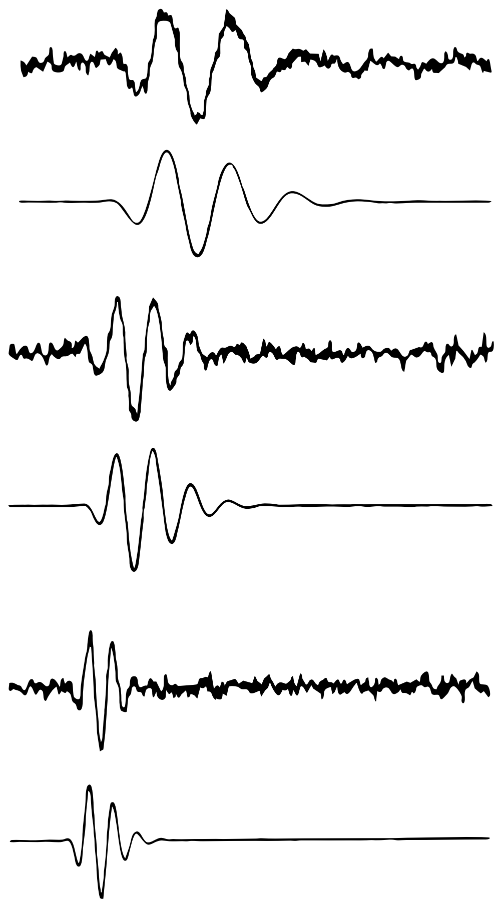
Auditory filters estimated by reverse correlation



Cat auditory “revcor” filters:

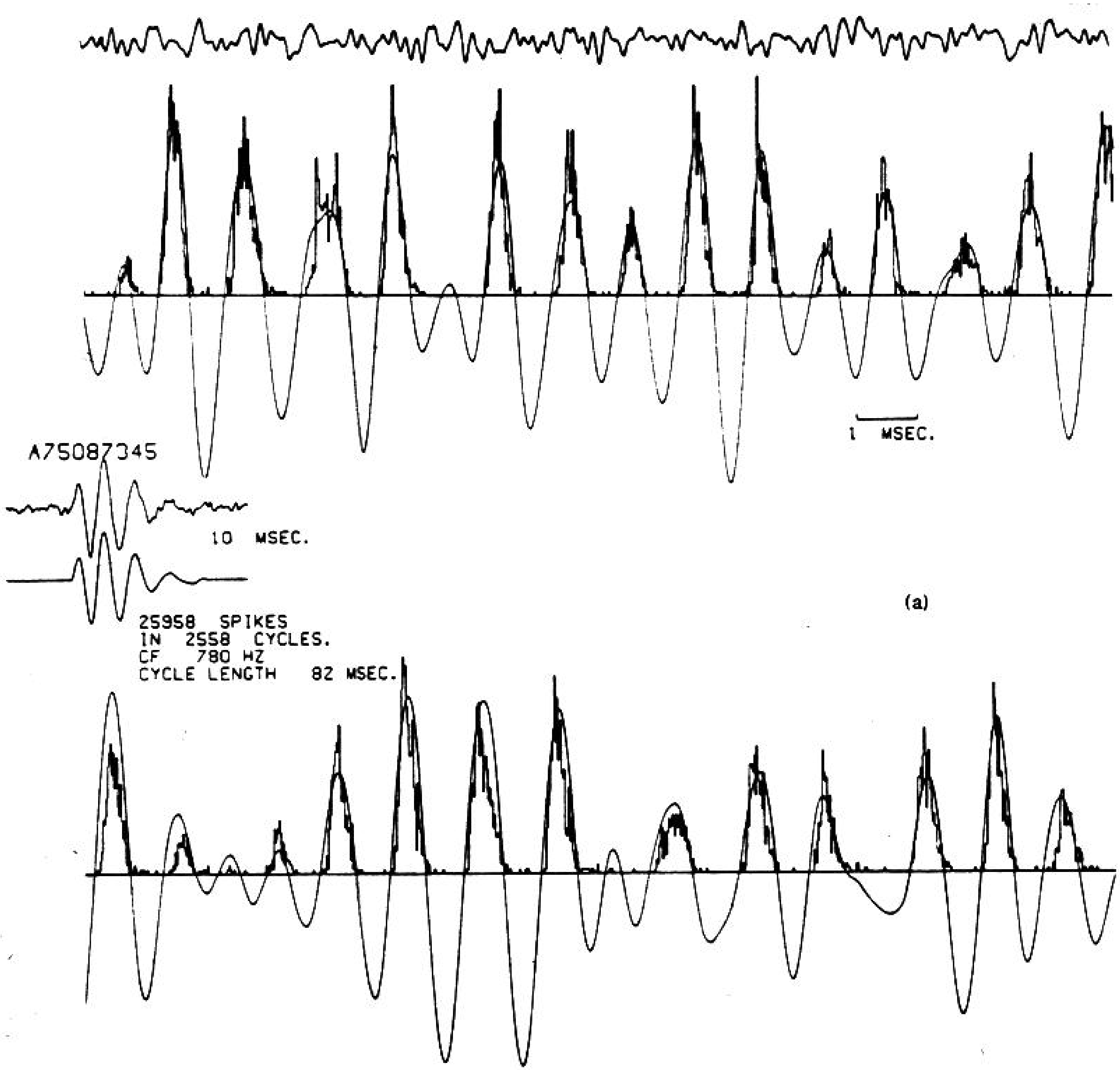


deBoer and deJongh, 1978



Carney and Yin, 1988

Revcor filter predictions of auditory nerve response



(from de Boer and de Jongh, 1978).

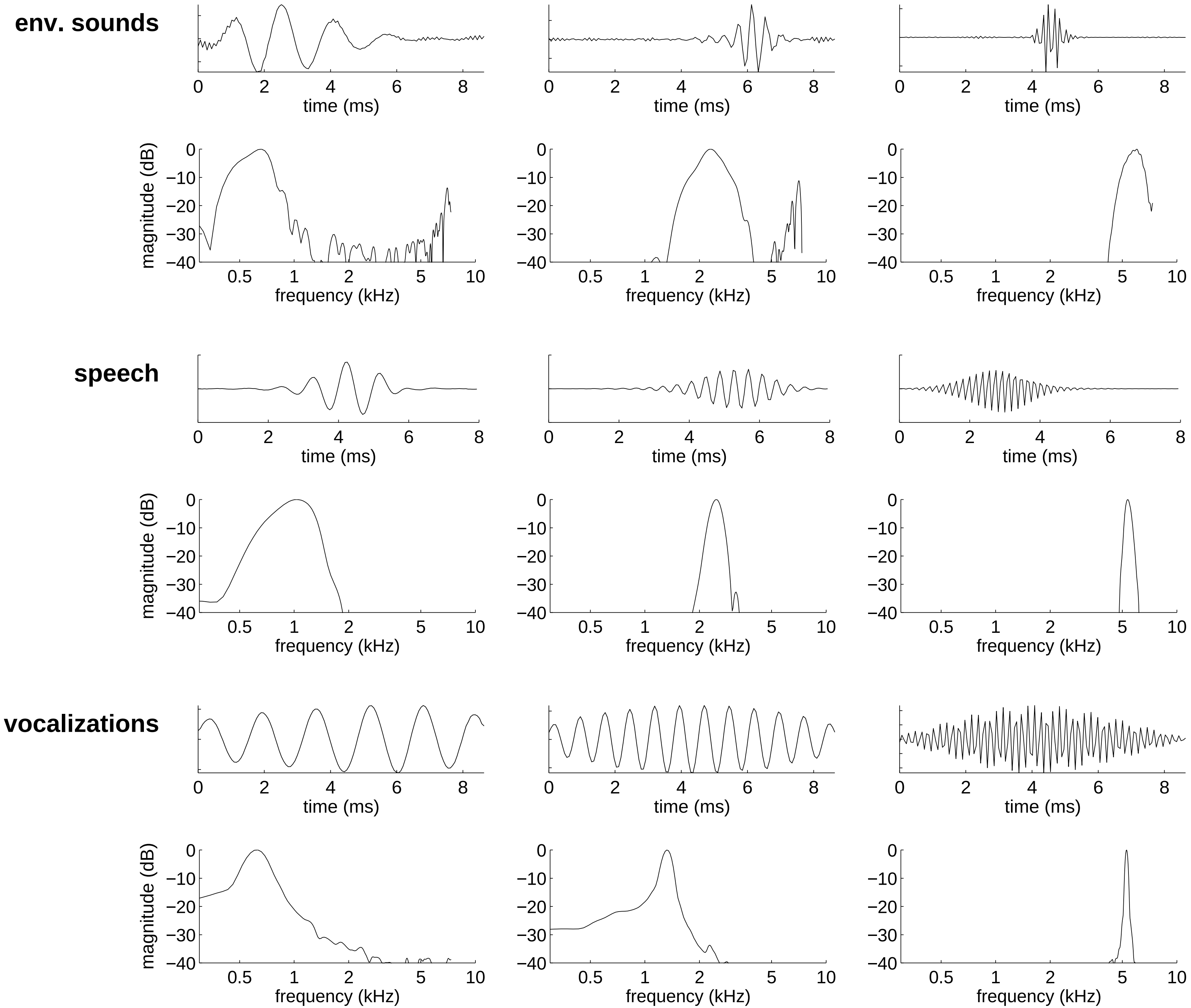
- stimulus is white noise
- histogram: measured auditory nerve response
- smooth curve: predicted response

Conclusion:

> *Shape and distribution of revcor filters account for a large part of the auditory sensory code.*

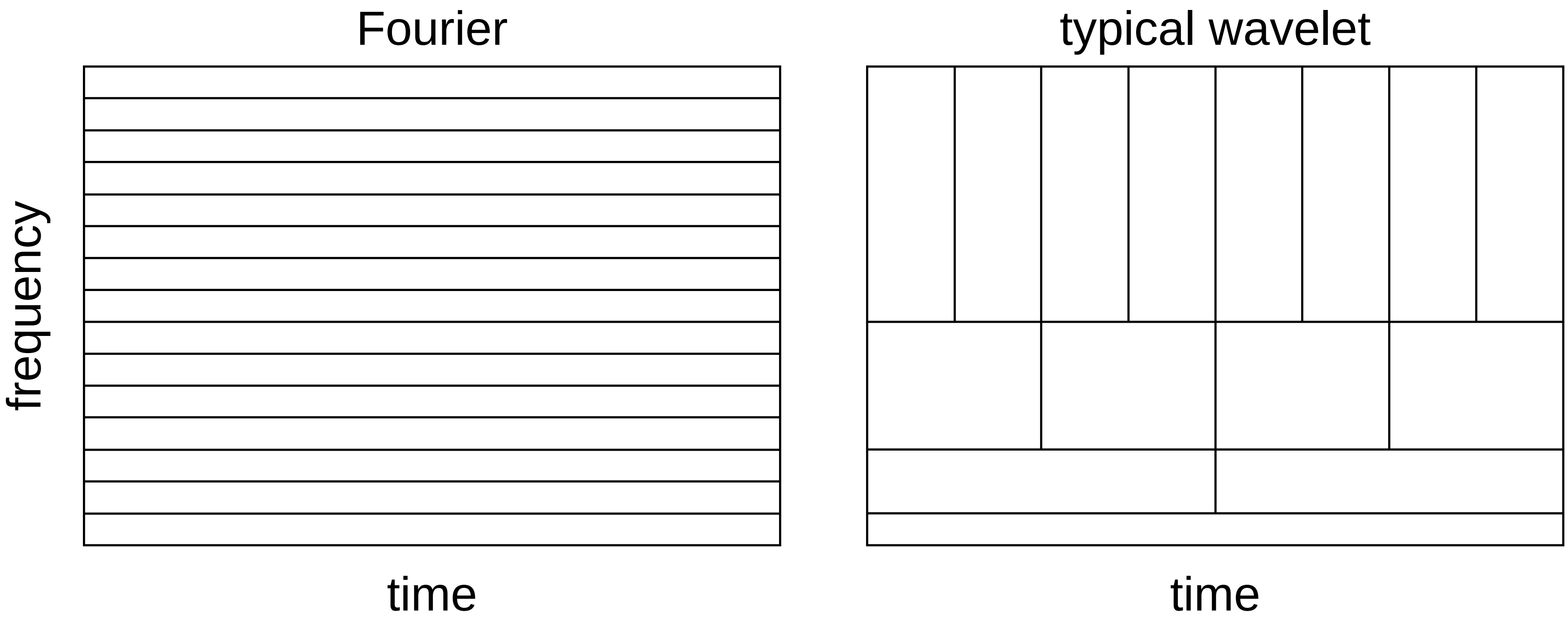
We want to match more than just individual filters:

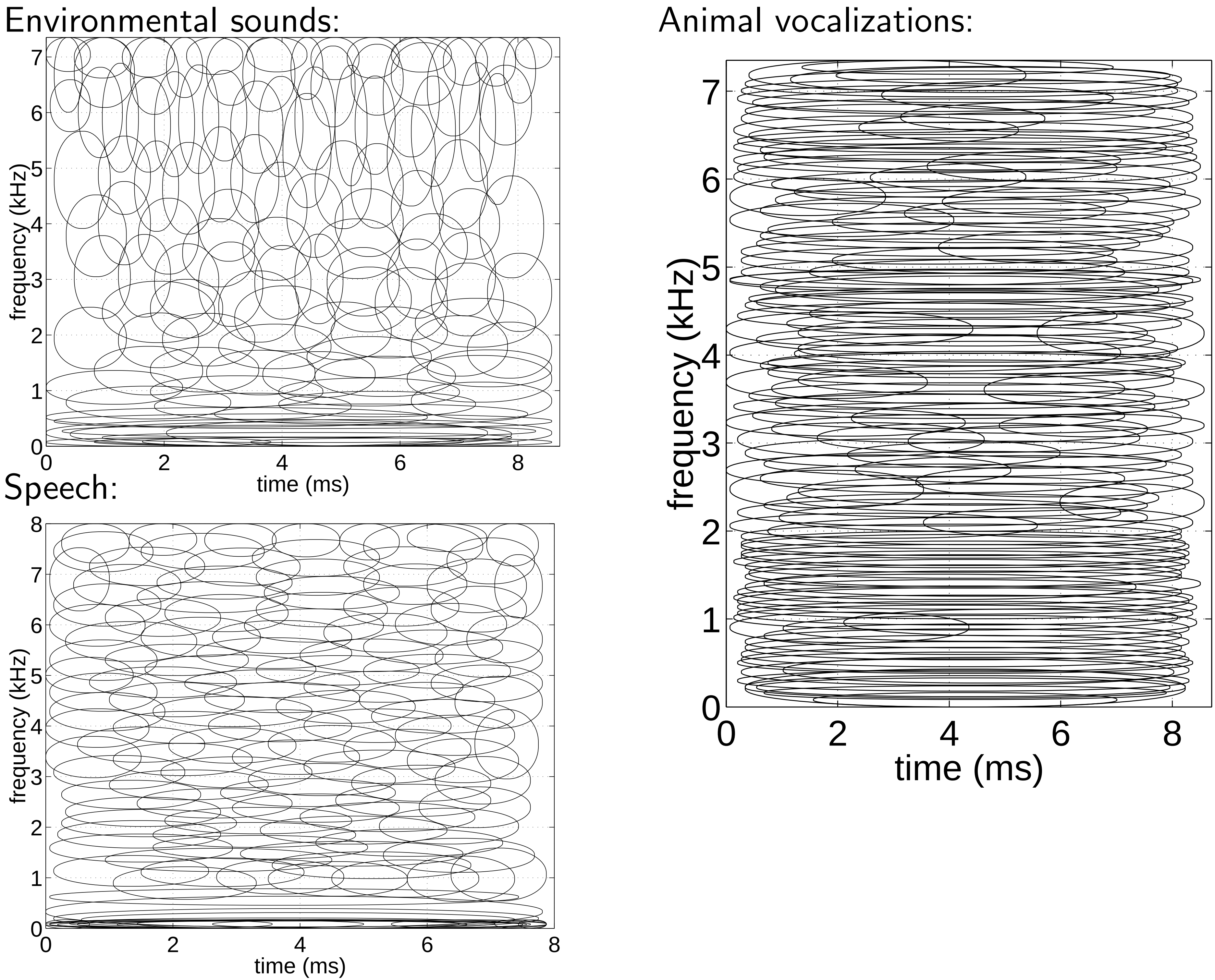
> *How do we characterize the population?*



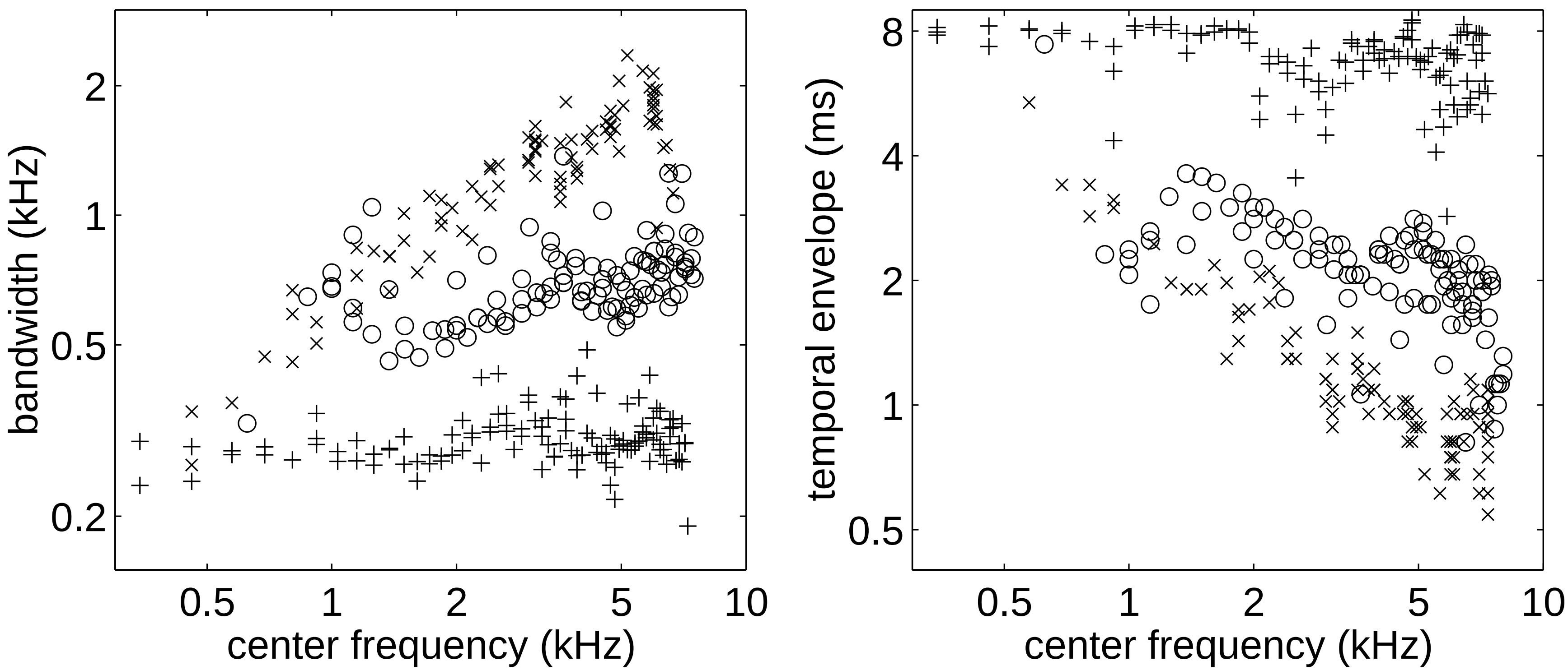
env. sounds
speech
vocalizations
time (ms)
frequency (kHz)
magnitude (dB)
0
2
4
6
8
0.5
1
2
5
10
0
−10
−20
−30
−40

Schematic time-frequency distributions




Environmental sounds:
frequency (kHz)
time (ms)
Speech:
frequency (kHz)
time (ms)
Animal vocalizations:
frequency (kHz)
time (ms)

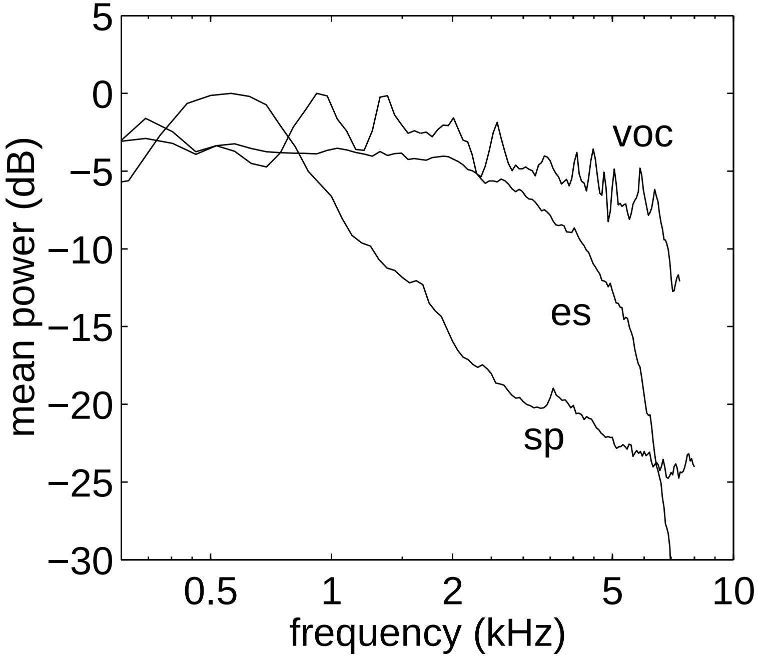
Tiling trends follow power law



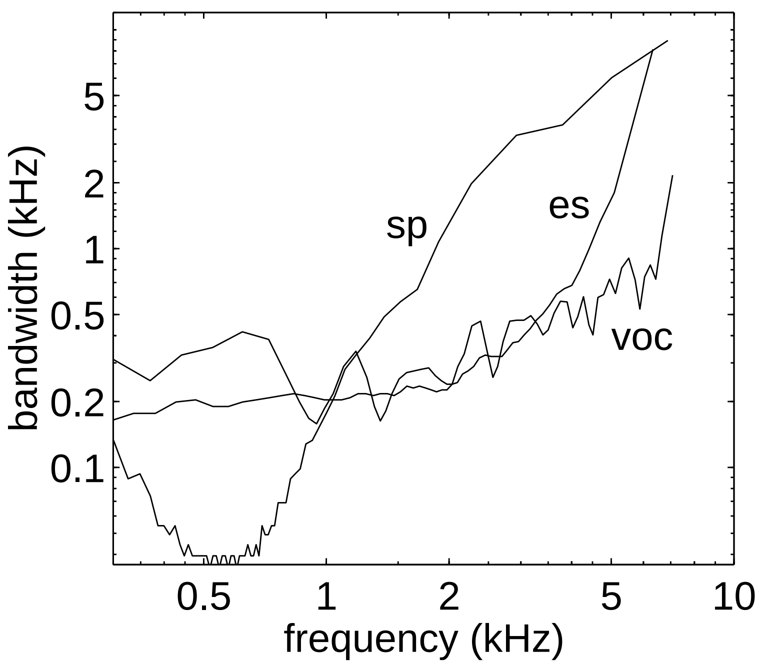
'×' = environmental sounds 'o' = speech '+' = vocalizations

Does equalization of power explain these data?

Average power spectra:

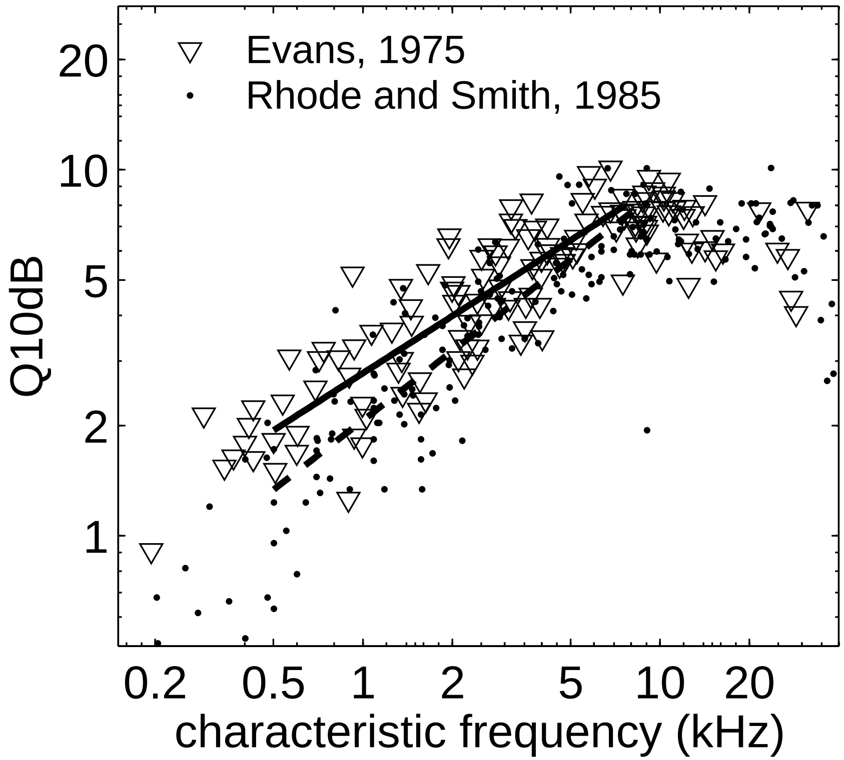


Equal power across frequency bands:

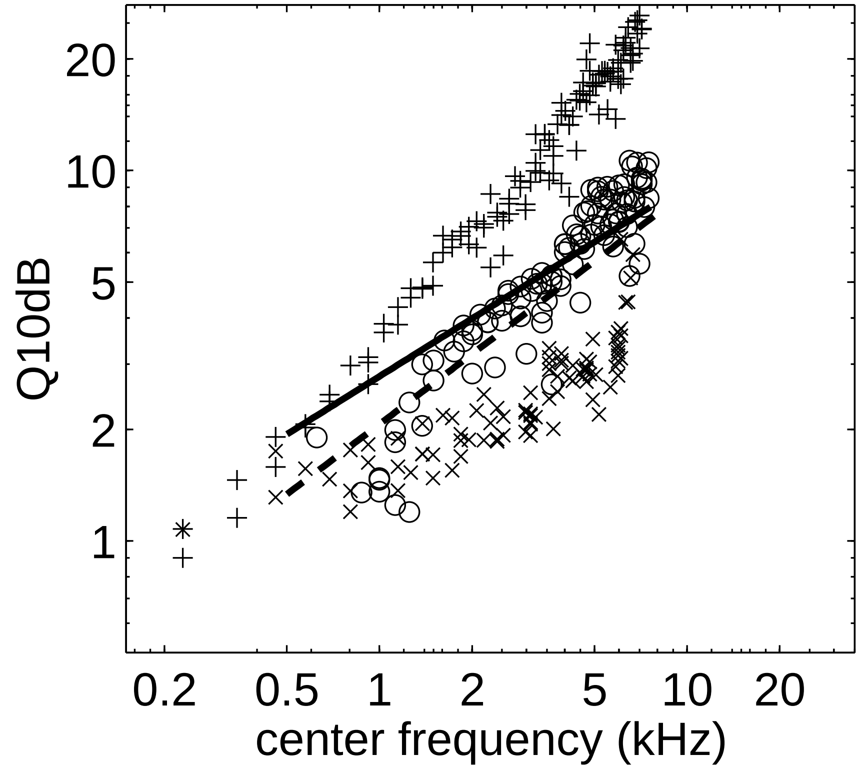


Comparison to auditory population code

Cat auditory nerves



Derived filters



Filter sharpness characterizes how bandwidth changes as a function of frequency

$$Q_{\mathrm{10dB}} = f_c / w_{10dB}$$

'+' vocalizations
'○' speech
'×' environmental sounds

Summary

Information theory and efficient coding:

- can be used to *derive* optimal codes for different pattern classes.
- explains important properties of sensory codes in both the auditory and visual system.
- gives insight into how our sensory systems are adapted to the natural environment.

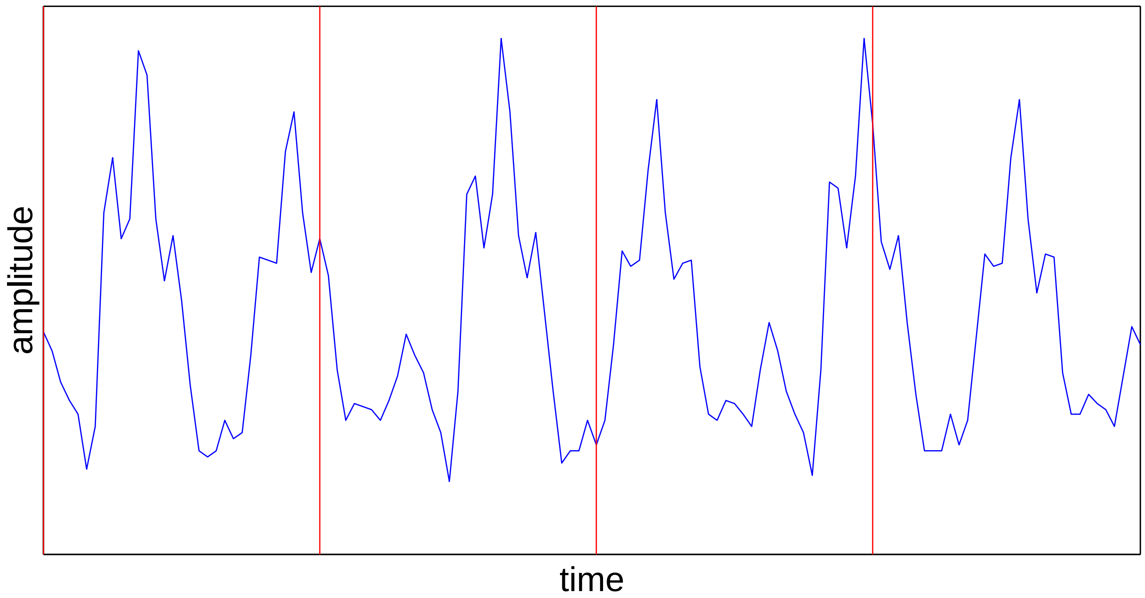
Caveats

- Codes can only be derived within a small window
- Does not explain non-linear aspects of coding
- Models do not capture higher order structure

Coding natural sounds with spikes

Addressing some limitations of the current theory

The current model assumes the sound waveform is dividing into blocks:



Problems with block coding:

- signal structure is arbitrarily aligned
- code depends on block alignment
- difficult to encode non-periodic structure, e.g. rapid onsets

An efficient, shift-invariant model

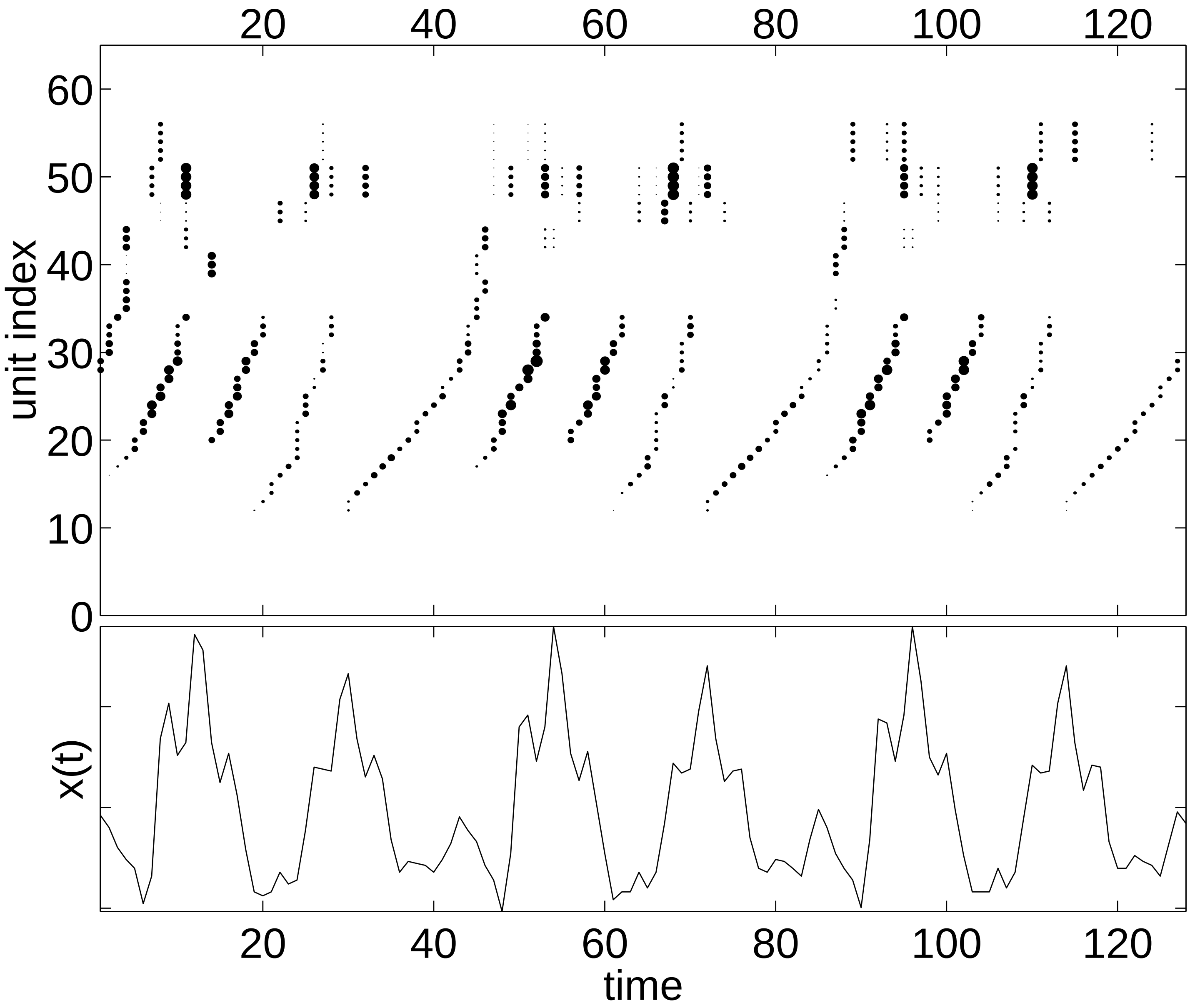
The signal is modeled by a sum of events plus noise:

$$x(t) = s_1\phi_1(t - \tau_1) + \cdots + s_M\phi_M(t - \tau_M) + \epsilon(t)\,.$$

The events $\phi_m(t)$:

- can be placed at arbitrary time points τ_m
- are scaled by coefficients s_m

Solution after optimization: 105 dB SNR



Time shifting