What are the problems of sensory coding?

- What should the sensor sense?
- How is energy transduced?
- How to deal with noise?
- How to compress dynamic range?
- How to prevent the sensor from being damaged?
Two approaches to the study of systems

1. Experimental/behavioral approach:
   - describe and characterize behavior
   - understand range and limitations
   - investigate system properties and organization
   - develop theories to better understand functional roles

2. Theoretical/computational approach:
   - define problem
   - develop models and algorithms
   - understand range and limitations
   - develop more algorithms: more general/specialized; faster/less resources
The complexity of the auditory system

from Warren, 1999
What principles should guide the choice of representation?

Unsupervised approaches:

• find useful “features”
• adapted to the patterns of interest
• useful in a wide range of tasks

Supervised approaches:

• Maximize performance on given task

At low-levels, we have to use unsupervised approaches.
Linear superposition

Goal is to describe the data to desired precision.
Code signal by linear superposition of basis functions:

\[
x = \tilde{a}_1 s_1 + \tilde{a}_2 s_2 + \cdots + \tilde{a}_L s_L + \tilde{\epsilon}
= A \hat{s} + \epsilon
\]

- \( x(t) \) is represented by a vector \( x \)
- \( \tilde{a}_i \) are the \textit{basis vectors}
- \( A \) is the \textit{basis} (could be Fourier, wavelet, etc.)
- \( s_i \) are the \textit{coefficients}

Can solve for \( \hat{s} \) in the no noise case

\[
\hat{s} = A^{-1} x
\]
An information theoretic approach

Want algorithm to choose optimal $A$ (basis matrix).

Generative model for data is:

$$x = As + \epsilon$$

Probability of pattern $x$ given representation $s$

$$P(x|A,s) \sim f(x - As, \Sigma, I)$$
Learning objective

Objective: maximize coding efficiency
⇒ maximize probability of data ensemble

Probability of pattern ensemble is:

\[ P(x_1, x_2, \ldots, x_N|A) = \prod_k P(x_k|A) \]
Optimal coding of an acoustic waveform

- We do not assume a Fourier or spectral representation.
- Goal:
  Predict optimal transformation of acoustic waveform from statistics of the acoustic environment.
- Use a simple model: bank of linear filters
Coding patterns with a statistical model

Goal: Encode the pattern to desired precision:

\[ x = \vec{a}_1 s_1 + \cdots + \vec{a}_L s_L + \vec{\epsilon} \]
\[ = \mathbf{A}s + \epsilon \]

Posterior:

\[ P(s|x, A) = \frac{P(s)P(x|s, A)}{P(x|A)} \]

Prior: \( s_i \)'s are independent and sparse:

\[ P(s) = \prod_i P(s_i) \]
\[ P(s_i) \propto \exp \left[ -\frac{|s_i|}{\lambda_i} \right] \]
Coding patterns with a statistical model

**Goal:** Encode the patterns to desired precision:

\[
x = \vec{a}_1 s_1 + \cdots + \vec{a}_L s_L + \vec{e}
\]

\[
= \mathbf{A}s + \epsilon
\]

**Posterior:**

\[
P(s|x, \mathbf{A}) = \frac{P(s)P(x|s, \mathbf{A})}{P(x|\mathbf{A})}
\]

**Prior:** \(s_i\)'s are independent and sparse:

\[
P(s) = \prod_i P(s_i)
\]

\[
P(s_i) \propto \exp \left[ - \left| \frac{s_i}{\lambda_i} \right|^{q_i} \right]
\]

- \(q=4\) (Gaussian)
- \(q=2\)
- \(q=1\) (Laplacian)
- \(q=0.5\)
Coding patterns with a statistical model

**Goal:** Encode the patterns to desired precision:

\[ x = \mathbf{a}_1 s_1 + \cdots + \mathbf{a}_L s_L + \mathbf{e} \]

\[ = \mathbf{A}s + \epsilon \]

**Posterior:**

\[ P(s|x, \mathbf{A}) = \frac{P(s)P(x|s, \mathbf{A})}{P(x|\mathbf{A})} \]

**Prior:** \( s_i \)'s are independent and sparse:

\[ P(s) = \prod_i P(s_i) \]

\[ P(s_i) \propto \exp\left[-\frac{|s_i|^q}{\lambda_i}\right] \]

**Likelihood:** Assume \( \epsilon \sim \text{Gaussian}, \)

\[ P(x|s, \Sigma) \propto \exp\left[-\frac{1}{2} \epsilon^T \Sigma^{-1} \epsilon \right] \]

**Inference:** use the MAP value:

\[ \hat{s} = \arg \max_s P(s|x, \mathbf{A}) \]

Simple special case: no noise (ICA)

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

**Inference (or recognition or coding):**

*finds most efficient representation of pattern \( x \) in a given basis \( \mathbf{A} \)*
Learning: Optimizing the model parameters

Learning objective:

\[
\text{maximize coding efficiency} \quad \Rightarrow \quad \text{maximize } P(x|A) \text{ over } A.
\]

Probability of pattern ensemble is:

\[
P(x_1, x_2, ..., x_N|A) = \prod_k P(x_k|A)
\]

Use \textit{independent component analysis} (ICA) to learn \( A \):

\[
\Delta A \propto A A^T \frac{\partial}{\partial A} \log P(x|A)
= -A(zs^T - I),
\]

where \( z = (\log P(s))' \). Assume generalized Gaussians:

\[
P(s_i) \sim \mathcal{N}^{q_i}(s_i|\mu, \sigma).
\]

This learning rule:

- \textit{learns the feature set that captures the most structure}
- \textit{optimizes basis to maximize the efficiency of the code}
Learning the optimal codes

Goal:

\[ \text{Predict optimal transformation of sound waveform from statistics of the acoustic environment} \]

Learning procedure:

- random sound segments (8 msec)
- optimize features using ICA

What sounds to use?

What tasks are auditory systems adapted to do?

- localization ⇒ environmental sounds
- communication ⇒ vocalizations
- general sound recognition

Use a variety of sound ensembles:

- non-harmonic \textit{environmental sounds} (e.g. footsteps, stream sounds, etc.)
- \textit{animal vocalizations} (rainforest mammals, e.g. chirps, screeches, cries, etc.)
- \textit{speech} (samples from 100 male & female speakers from the TIMIT corpus)
## Natural sounds

<table>
<thead>
<tr>
<th>Vocalizations</th>
<th>Environmental sounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>fox</td>
<td>walking on leaves</td>
</tr>
<tr>
<td></td>
<td>rustling leaves</td>
</tr>
<tr>
<td>squirrel</td>
<td>cracking branches</td>
</tr>
<tr>
<td></td>
<td>stream by waterfall</td>
</tr>
</tbody>
</table>
## Natural sounds

<table>
<thead>
<tr>
<th>Vocalizations</th>
<th>Environmental sounds</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>fox</strong></td>
<td>walking on leaves</td>
</tr>
<tr>
<td></td>
<td>rustling leaves</td>
</tr>
<tr>
<td>squirrel</td>
<td>cracking branches</td>
</tr>
<tr>
<td></td>
<td>stream by waterfall</td>
</tr>
</tbody>
</table>
### Natural sounds

<table>
<thead>
<tr>
<th>vocalizations</th>
<th>environmental sounds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>transient</td>
</tr>
<tr>
<td>fox</td>
<td>walking on leaves</td>
</tr>
<tr>
<td>squirrel</td>
<td>cracking branches</td>
</tr>
</tbody>
</table>
## Natural sounds

<table>
<thead>
<tr>
<th>vocalizations</th>
<th>environmental sounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>fox</td>
<td>walking on leaves</td>
</tr>
<tr>
<td>squirrel</td>
<td>cracking branches</td>
</tr>
<tr>
<td></td>
<td>rustling leaves</td>
</tr>
<tr>
<td></td>
<td>stream by waterfall</td>
</tr>
<tr>
<td></td>
<td>ambient</td>
</tr>
</tbody>
</table>
Natural sounds

<table>
<thead>
<tr>
<th>vocalizations</th>
<th>environmental sounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>fox</td>
<td>transient</td>
</tr>
<tr>
<td></td>
<td>walking on leaves</td>
</tr>
<tr>
<td></td>
<td>rustling leaves</td>
</tr>
<tr>
<td>squirrel</td>
<td>transient</td>
</tr>
<tr>
<td></td>
<td>cracking branches</td>
</tr>
<tr>
<td></td>
<td>stream by waterfall</td>
</tr>
<tr>
<td></td>
<td>ambient</td>
</tr>
<tr>
<td></td>
<td>rustling leaves</td>
</tr>
</tbody>
</table>
## Natural sounds

<table>
<thead>
<tr>
<th>Vocalizations</th>
<th>Environmental Sounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>fox</td>
<td>walking on leaves</td>
</tr>
<tr>
<td>squirrel</td>
<td>cracking branches</td>
</tr>
<tr>
<td></td>
<td>rustling leaves</td>
</tr>
<tr>
<td></td>
<td>stream by waterfall</td>
</tr>
</tbody>
</table>
Natural sounds

<table>
<thead>
<tr>
<th>vocalizations</th>
<th>environmental sounds</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>fox</strong></td>
<td>walking on leaves</td>
</tr>
<tr>
<td></td>
<td>rustling leaves</td>
</tr>
<tr>
<td><strong>squirrel</strong></td>
<td>cracking branches</td>
</tr>
<tr>
<td></td>
<td><strong>stream by waterfall</strong></td>
</tr>
</tbody>
</table>

- *fox* vocalizations include walking on leaves and rustling leaves.
- *squirrel* vocalizations include cracking branches and stream by waterfall.
Optimal linear filters for natural sounds

The optimal code depends on the class of sounds being encoded:
- a wavelet-like transform is best for environmental sounds
- a Fourier-like transform is best for vocalizations
- an intermediate transform is best for speech or general natural sounds
Characterizing the filter population

time-frequency distributions

![Time-frequency distributions](image-url)
Schematic time-frequency distributions

Fourier

frequency

time

typical wavelet

time
Comparison to cat auditory nerve data

Filter sharpness:

\[ Q_{10dB} = \frac{f_c}{w_{10dB}} \]
Next time:
non-linear coding