# Making a Robotic Dog See and Hear

Daniel D. Lee

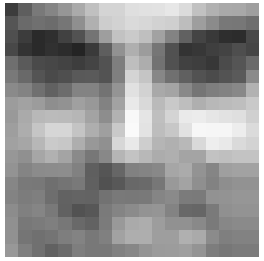
Lucent Technologies
Bell Labs Innovations



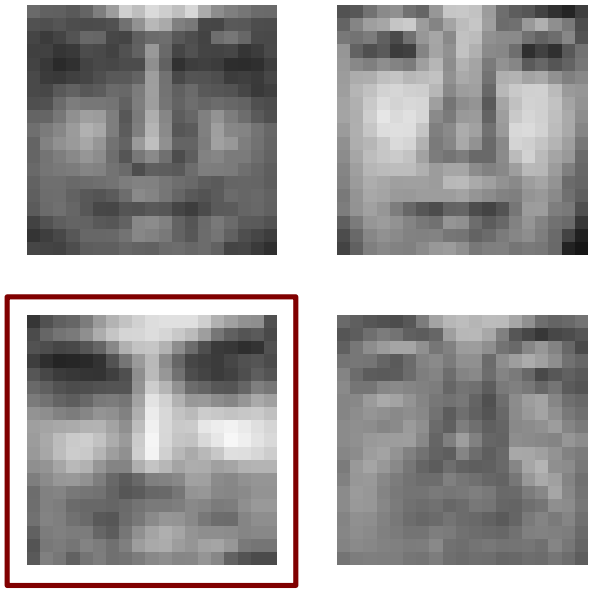
World of Science 2000

# Face recognition

Original image



Alternative images



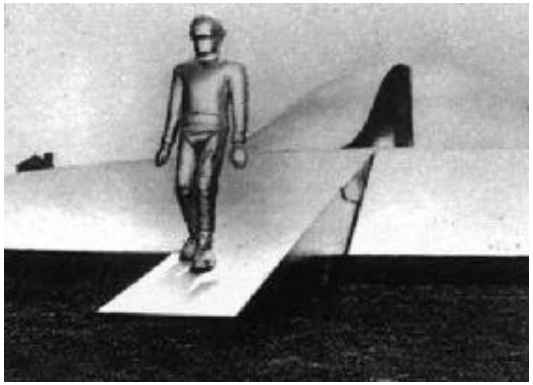
# Terminator





- Arnold is looking for you...

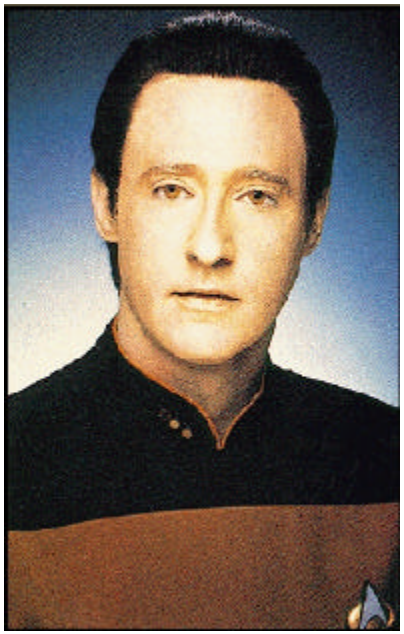
# Robots



Gort



HAL

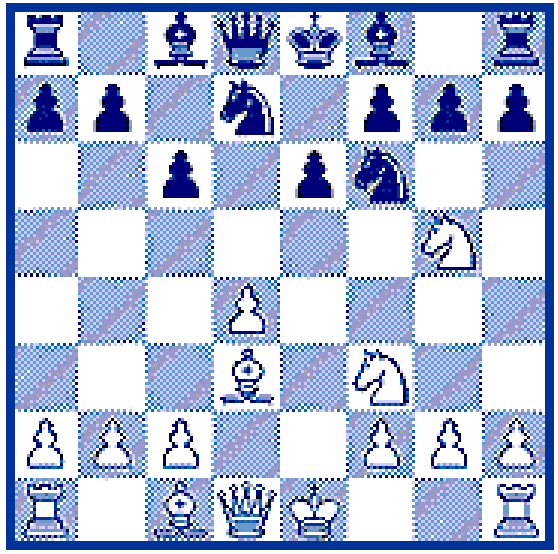


Data

- Hollywood versus reality

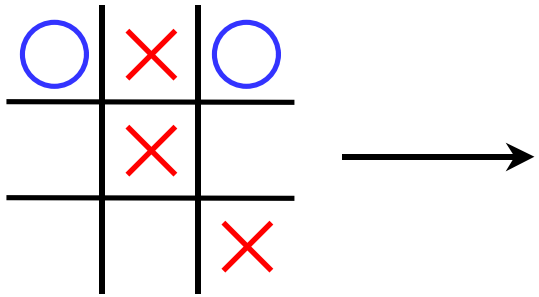
# Deep Blue





- Computer beats world champion Gary Kasparov

# Complexity



Tic Tac Toe

$$\begin{bmatrix} +1 \\ 0 \\ 0 \\ -1 \\ -1 \\ 0 \\ +1 \\ 0 \\ -1 \end{bmatrix}$$

Number of configurations

$$3^9 = 19683$$

- Tic Tac Toe easy to program using brute force
- Deep Blue evaluated 200 million chess positions per second

# Images

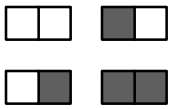


$$\begin{bmatrix} 0.0 \\ 0.5 \\ 0.7 \\ 1.0 \\ \star \\ 0.8 \\ 0.2 \\ 0.0 \end{bmatrix}$$

Pixel vector

- Vector representation of pixel values (white=0.0, black=1.0).

# Combinatorial explosion

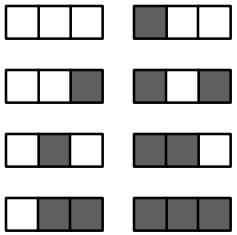


2 pixels

$2^2 = 4$ images



400 pixels



3 pixels

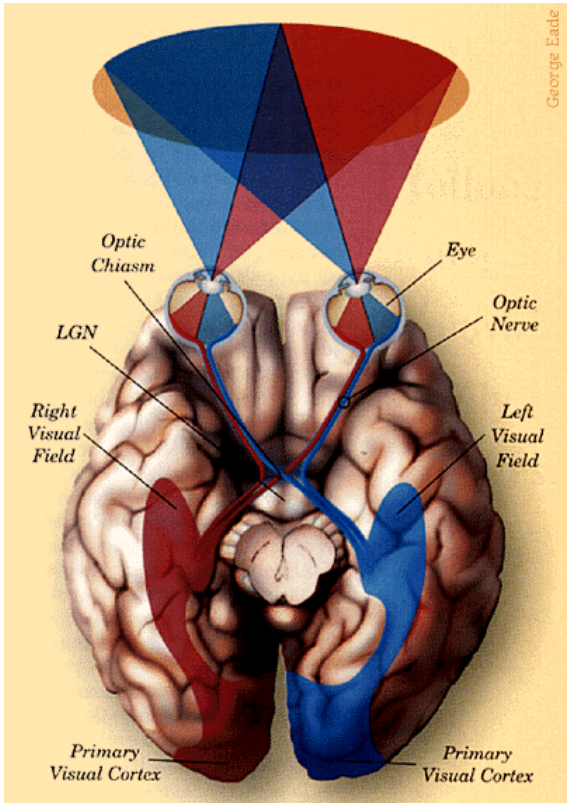
$2^3 = 8$ images

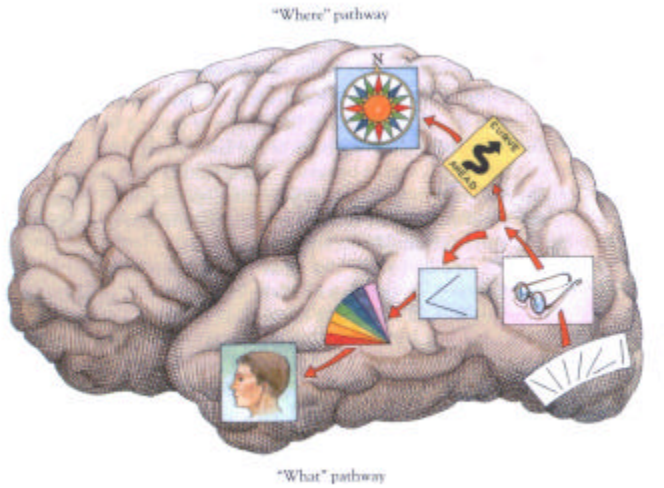
$2^{400} \approx 3 \times 10^{120}$ images

Age of universe:
$\approx 10^{17}$ seconds

- Impossible for a computer to search all possible images

# The brain

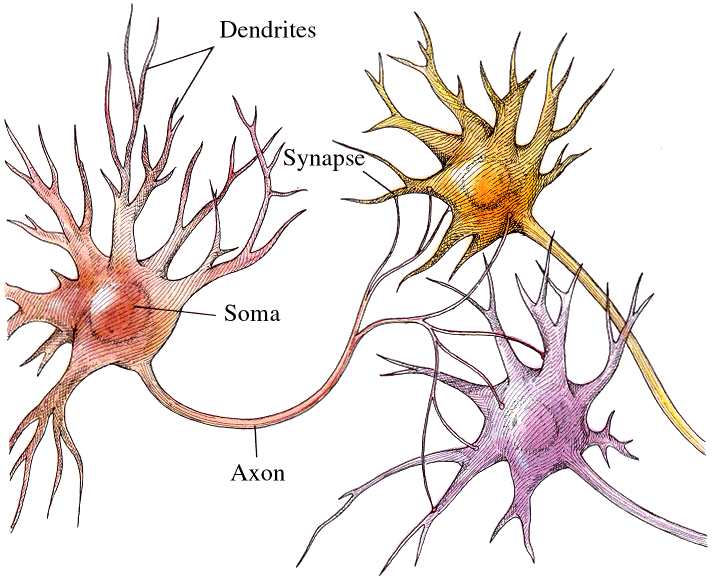



- Vision occupies a large fraction of our brains

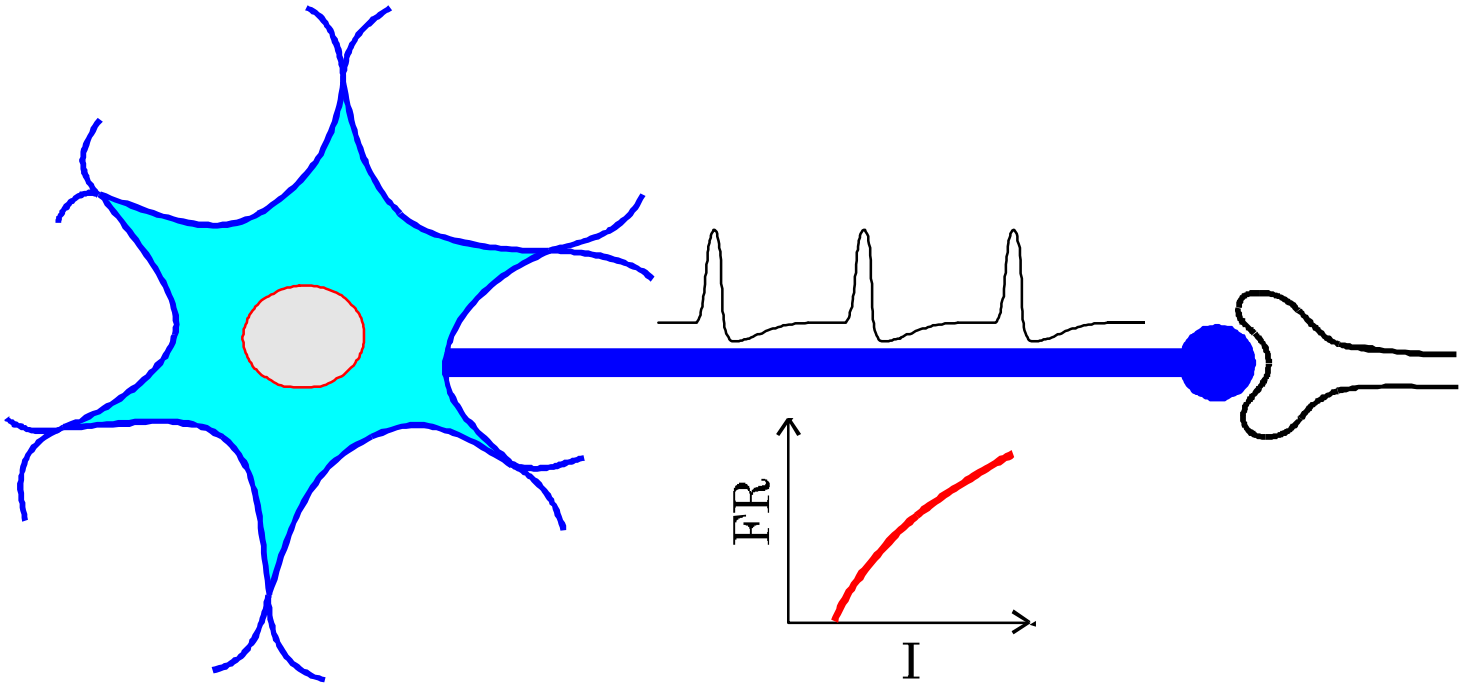
# Neurons





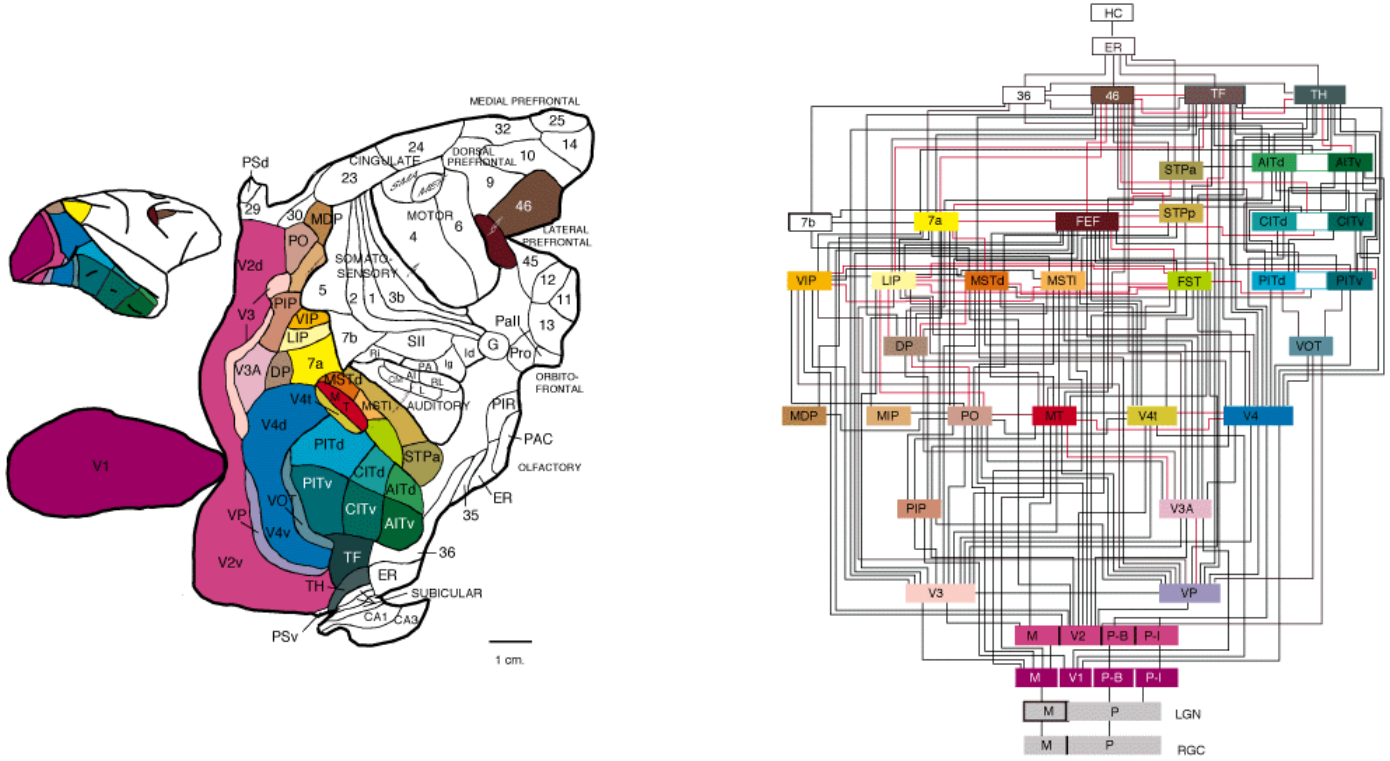

- Approximately $10^{12}$ neurons in a human brain

# Neuronal properties




- Neurons communicate with each other using action potentials

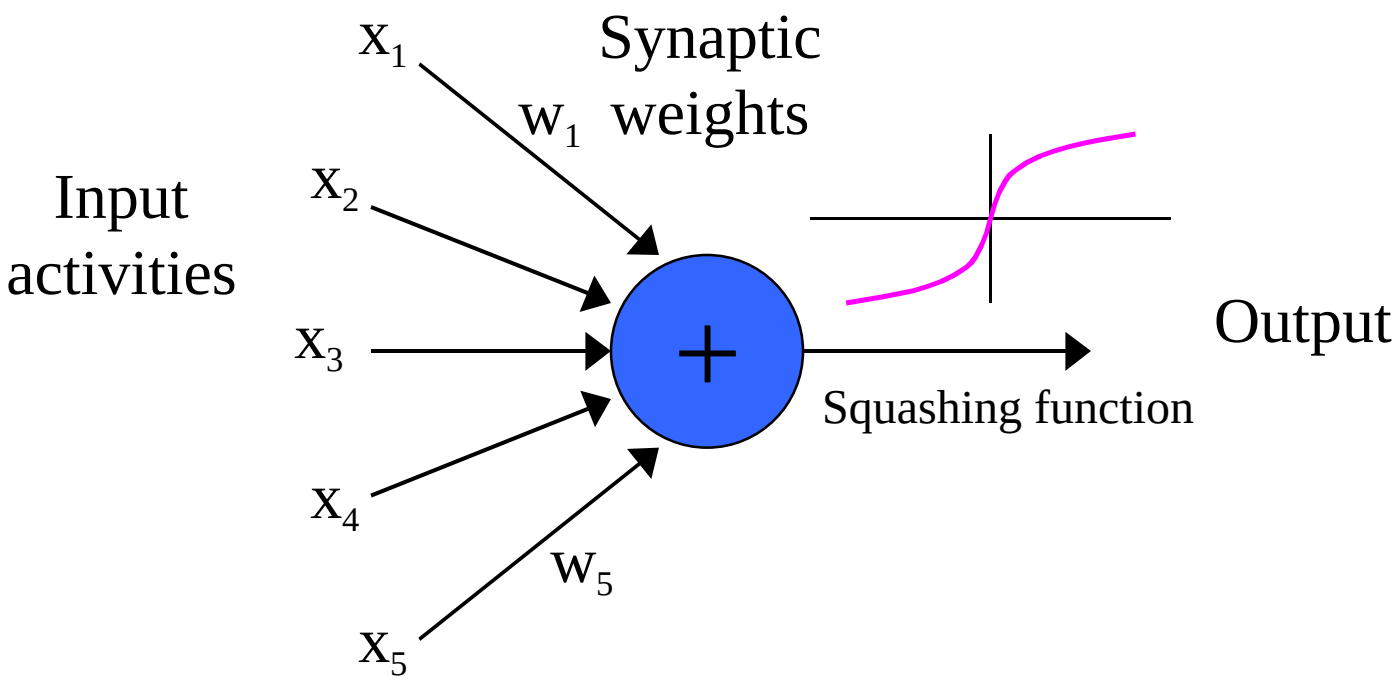
# Circuit diagram




(Felleman & Van Essen, 1991)

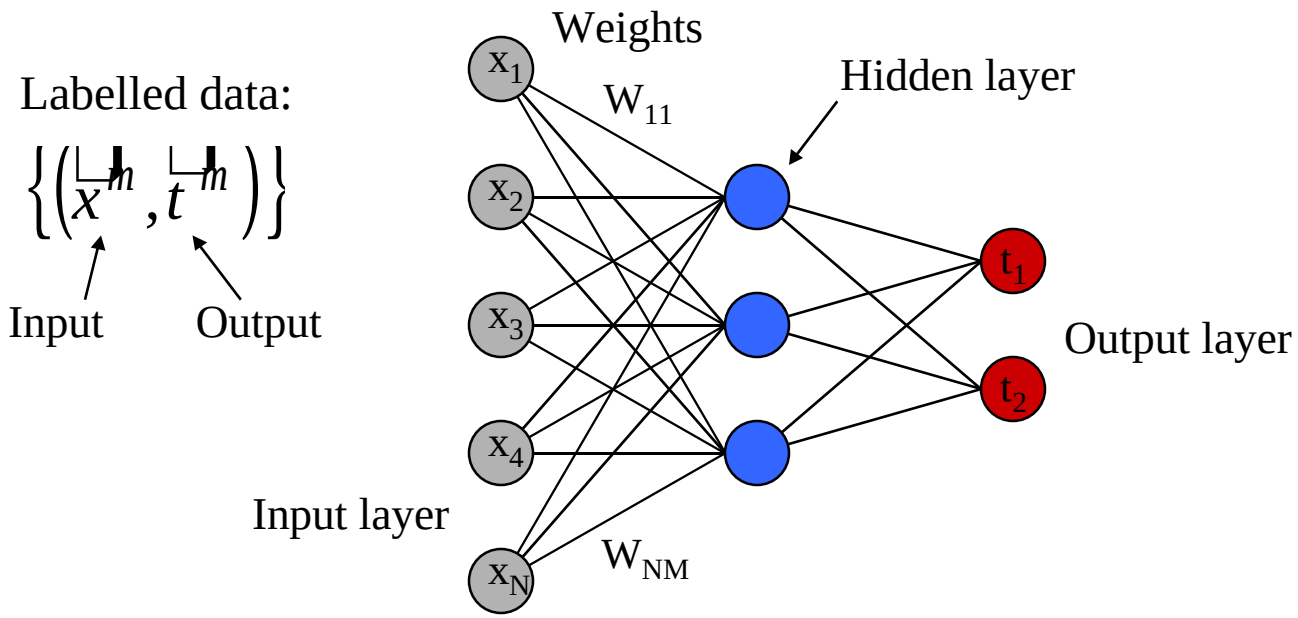
- Complex and hierarchical organization.

# Artificial neuron



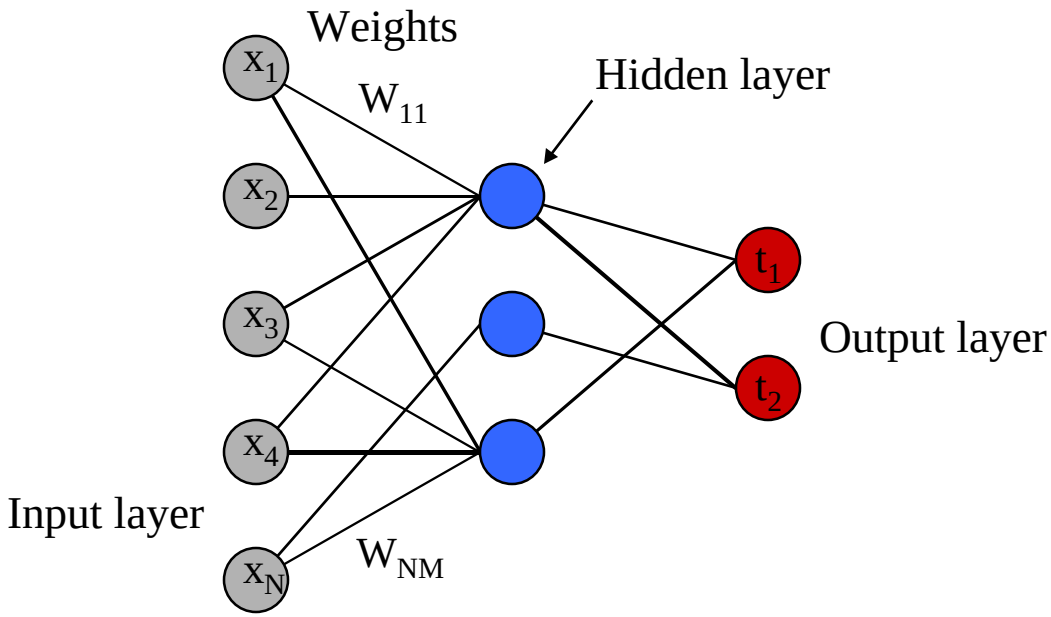

- Unit sums inputs **x** with synaptic weights **w**
- Nonlinear transformation

# Artificial neural network




- Transformation of input into output.
- Change synaptic weights to maximize performance.

# Learning

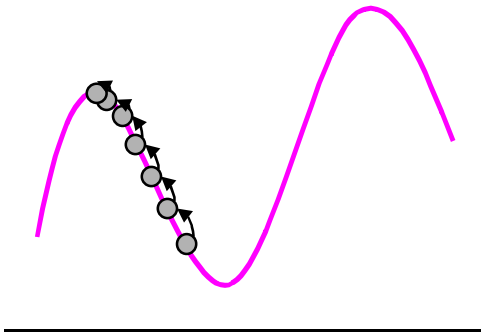



- How to set the connections between neurons to have the network do the right thing?

# Optimization



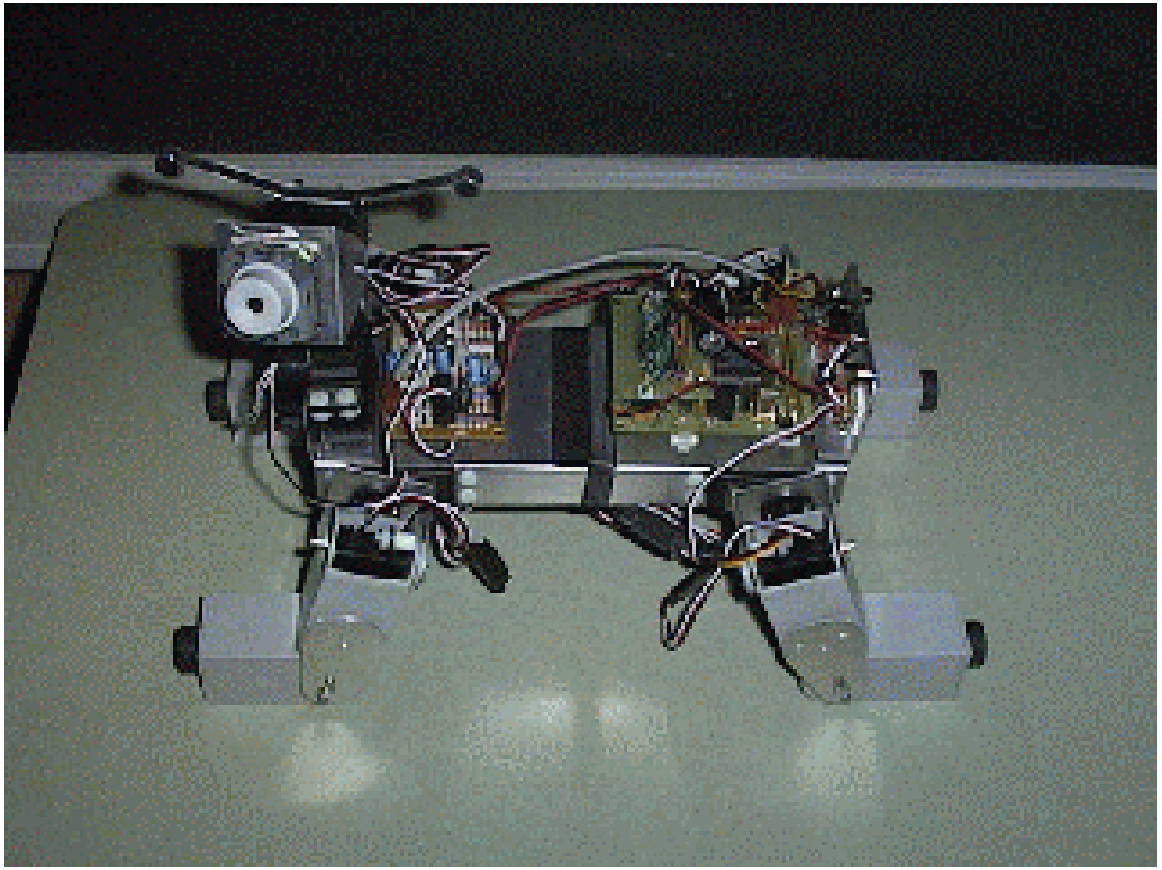
Mount Everest



Gradient ascent

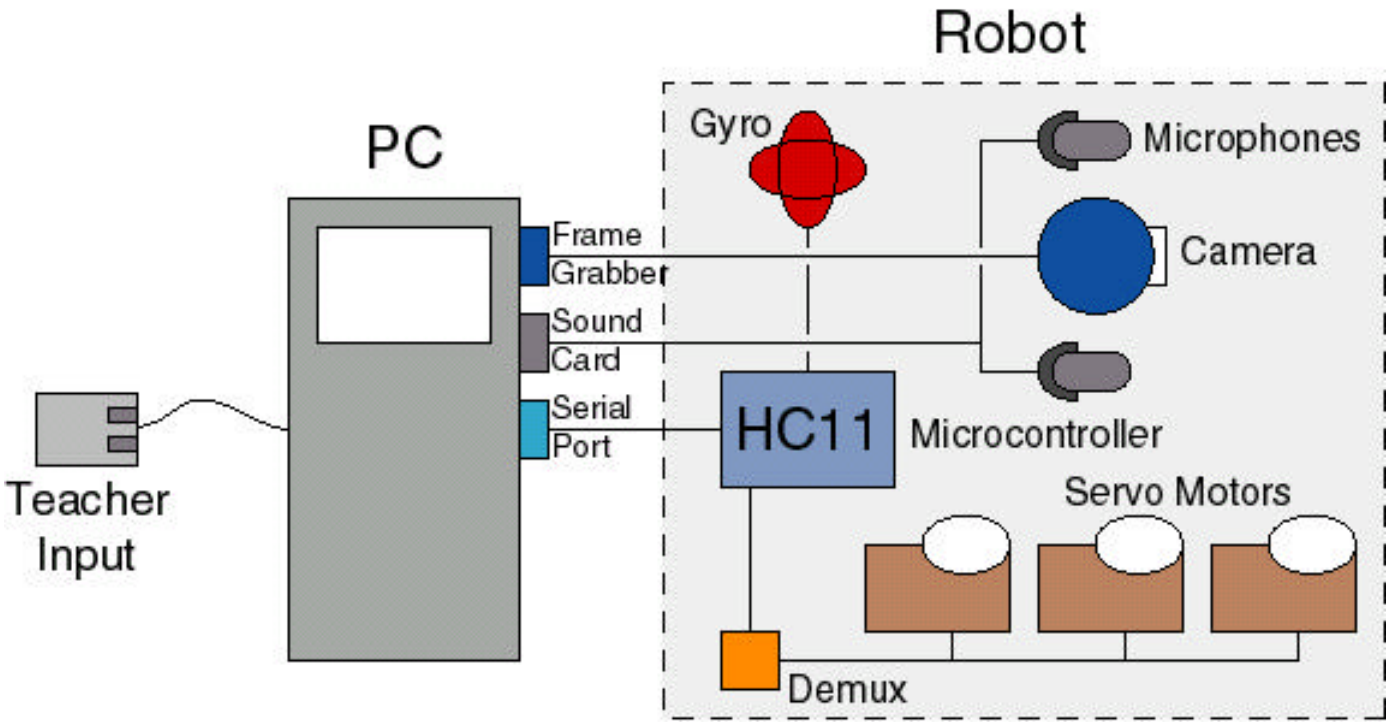
- Like climbing a mountain blindfolded.
- Small steps until top is reached.

# Robotic dog



- Doesn’t have a name yet… any suggestions?

# Artificial sensorimotor system

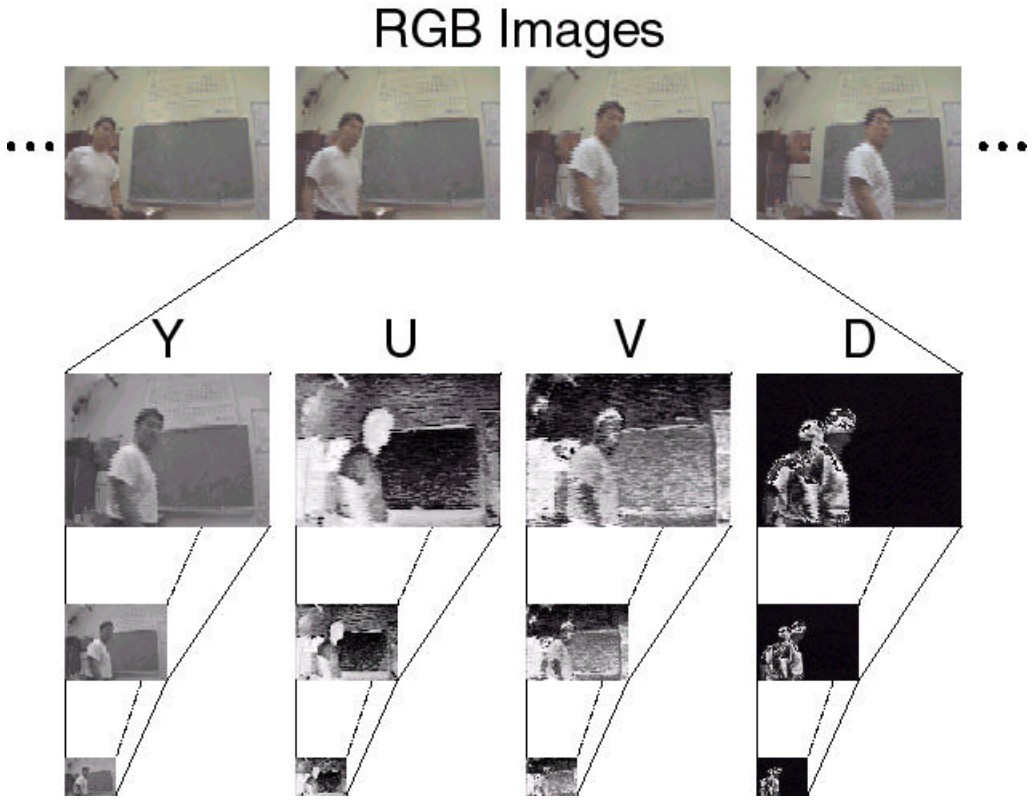



- Total cost of parts ~ $700
- You too can build your own!

# Video tracking

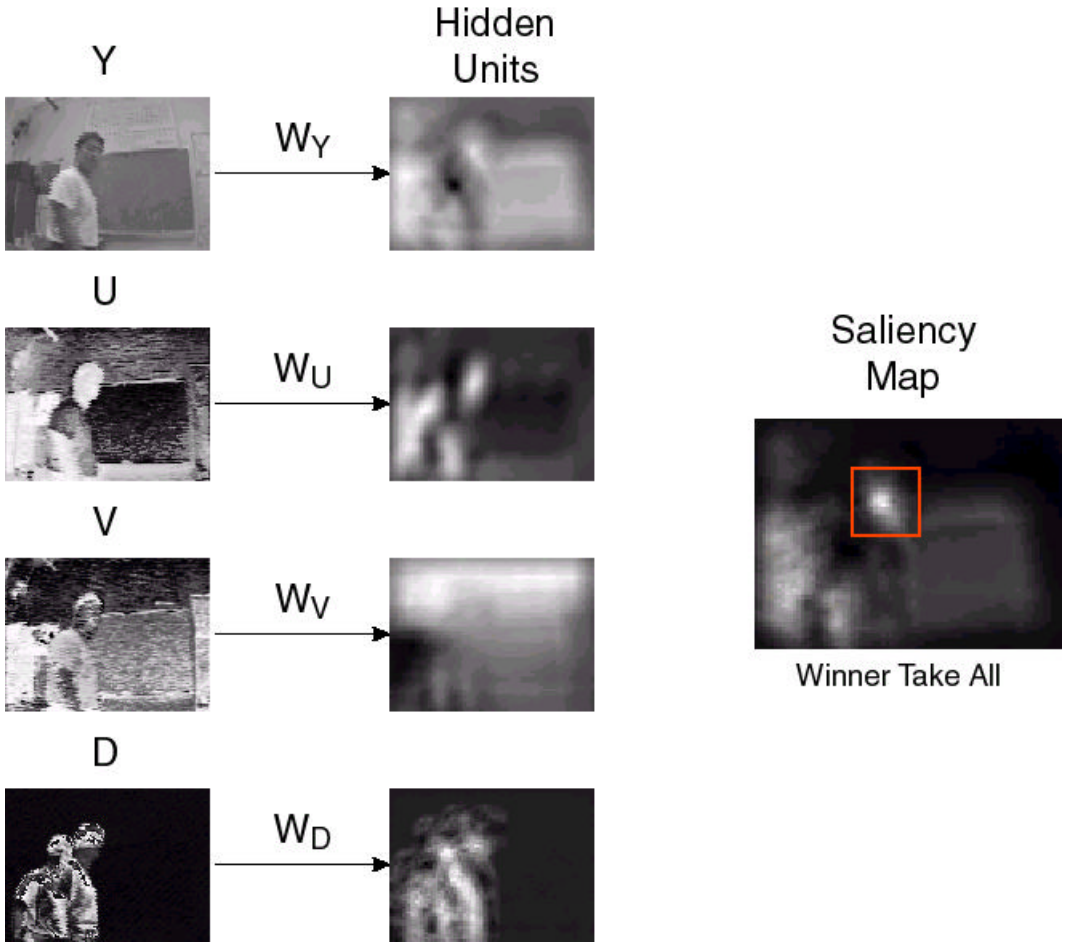


# Video processing




- Conversion of video images into luminance, color, and motion channels.

# Face recognition neural network



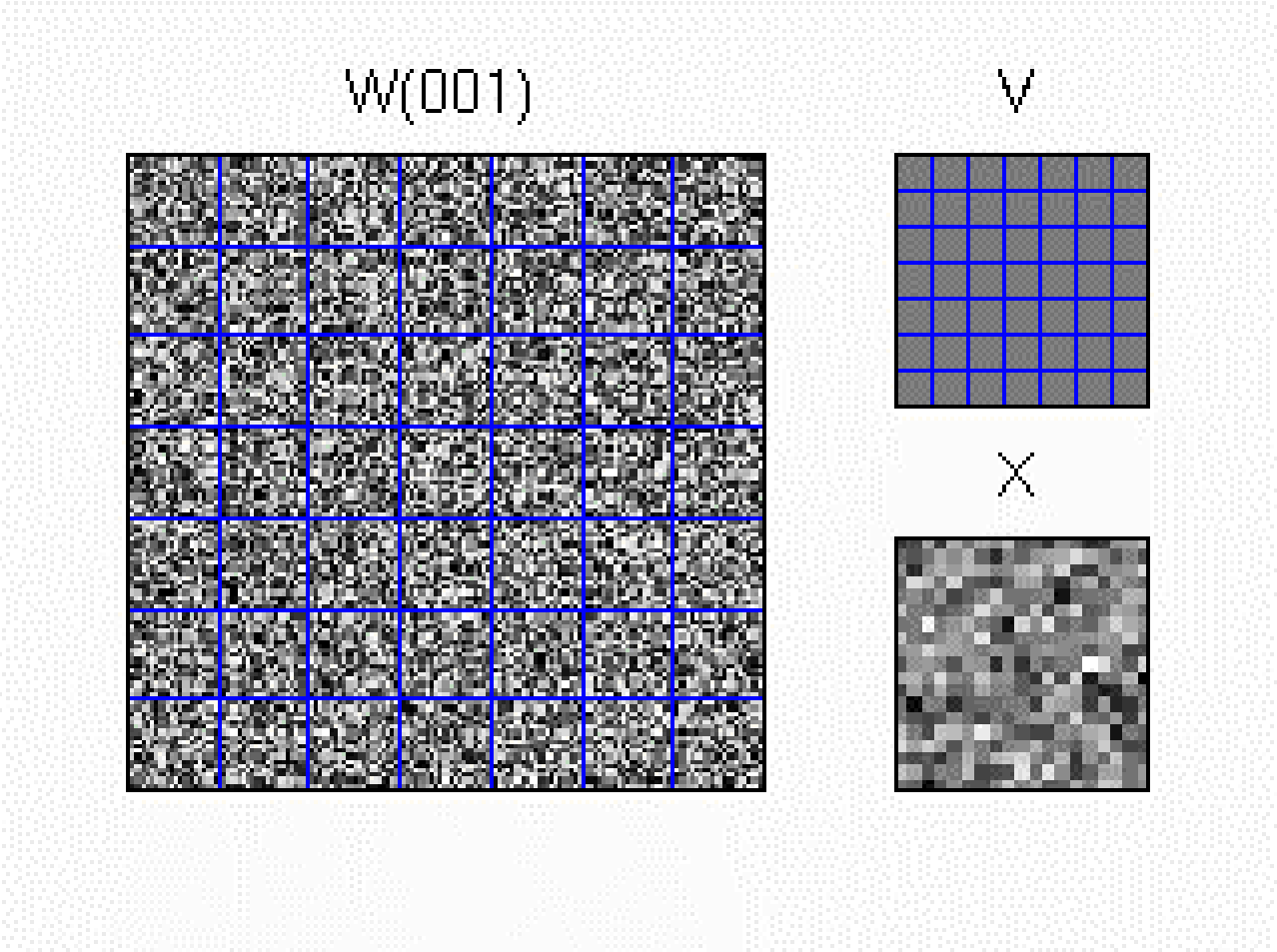

- Learns to associate saliency with face.

# Unsupervised learning

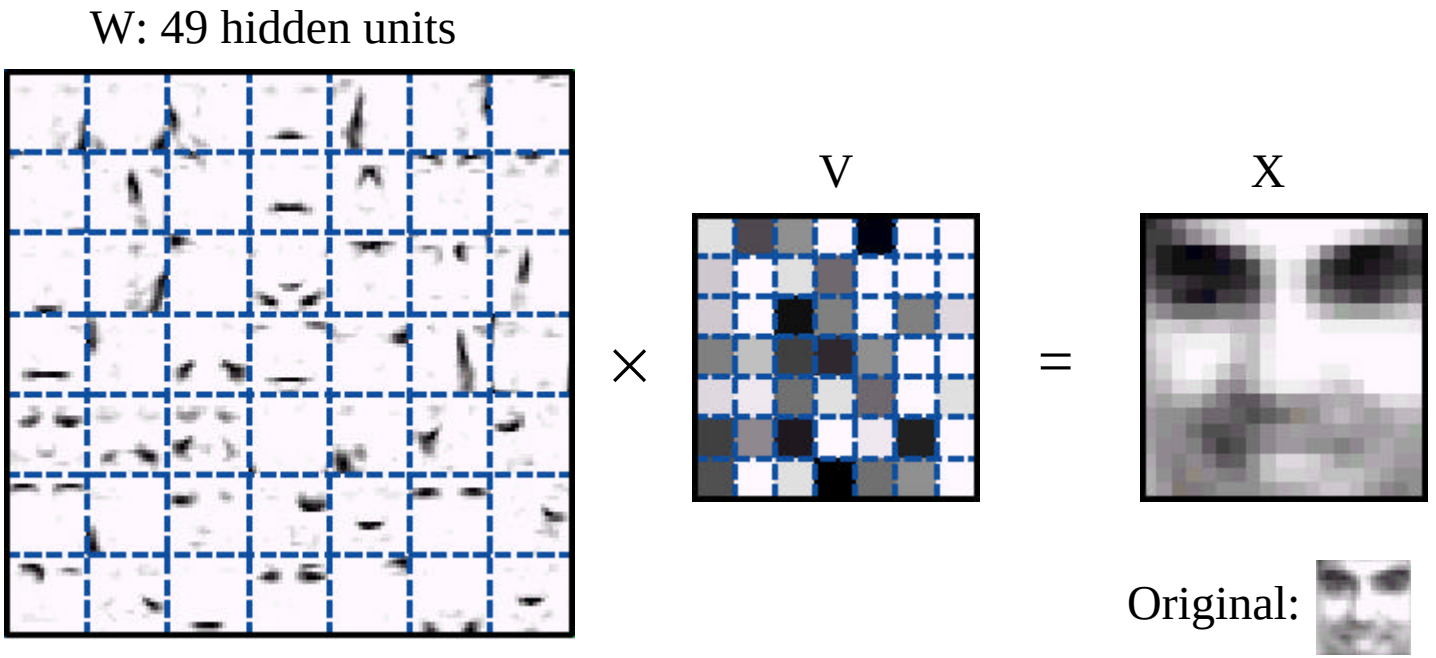


- Database containing many different faces.

# Learning parts of faces

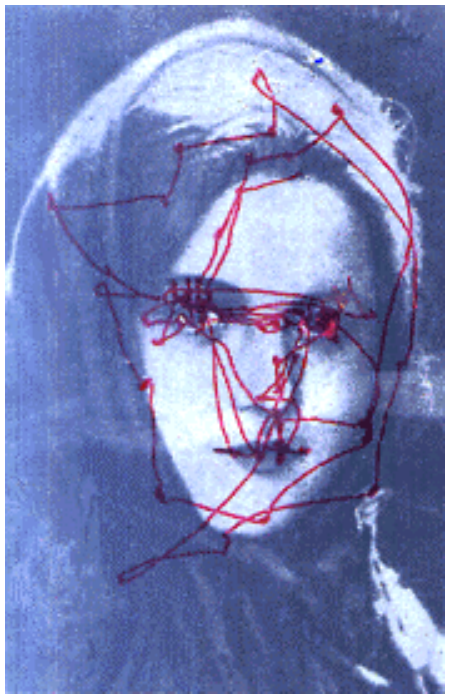


# Parts representation

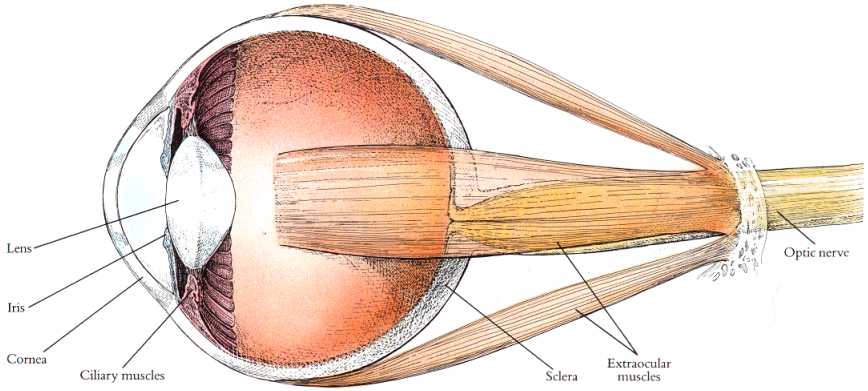



- Computer automatically decomposes the images into their constituent parts.

# Eye movements



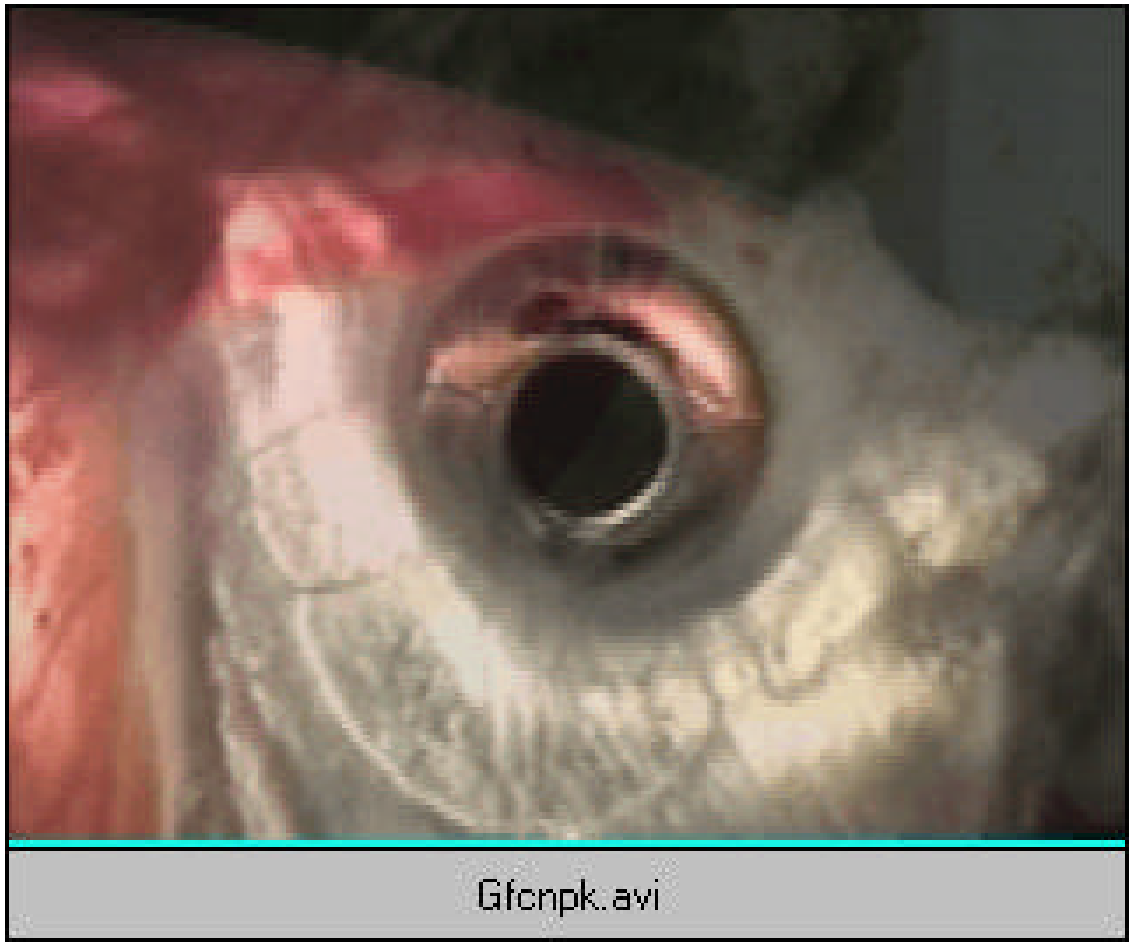
(Yarbus, 1967)



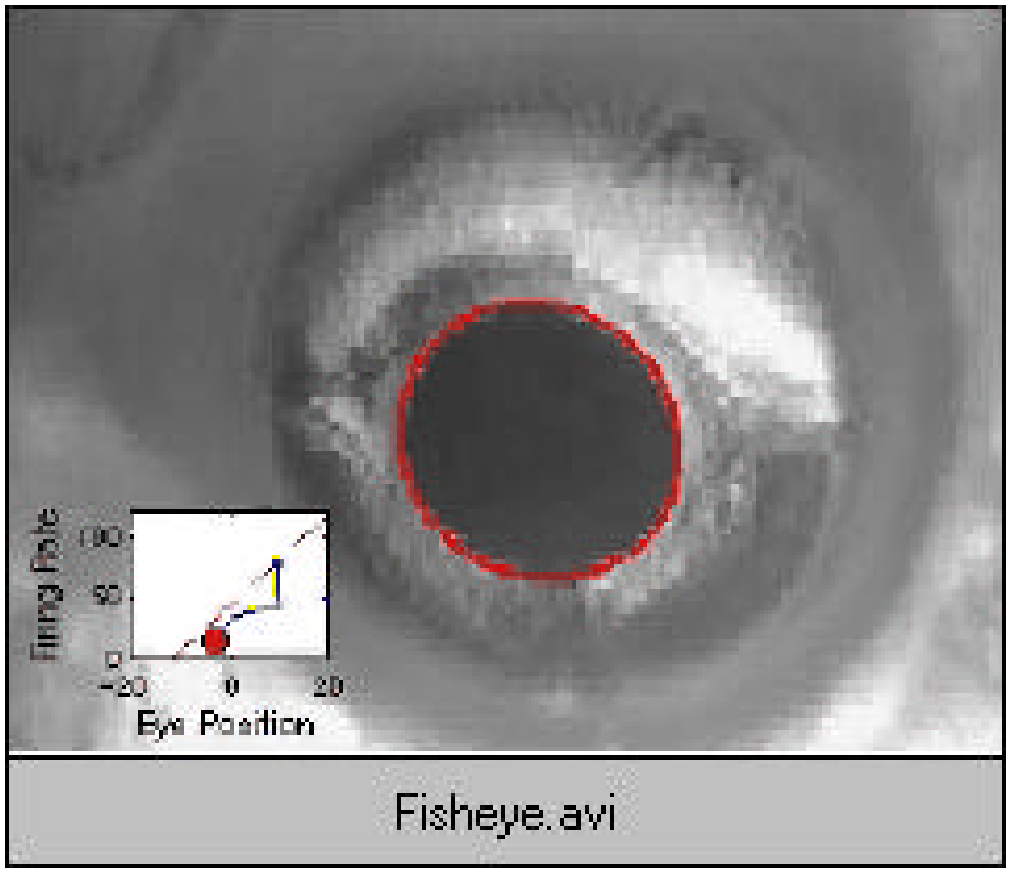

Eye muscles

- Fast eye movements to scan visual environment

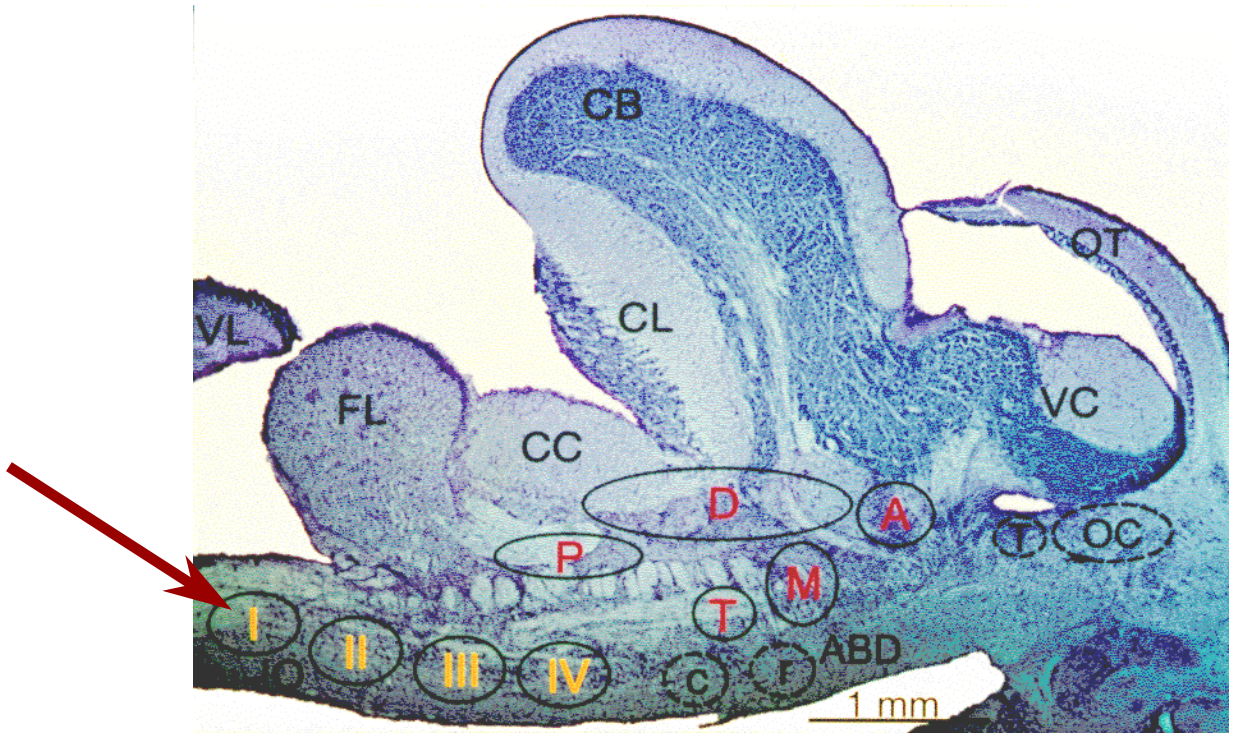
# Goldfish eye movements



# Control of eye position

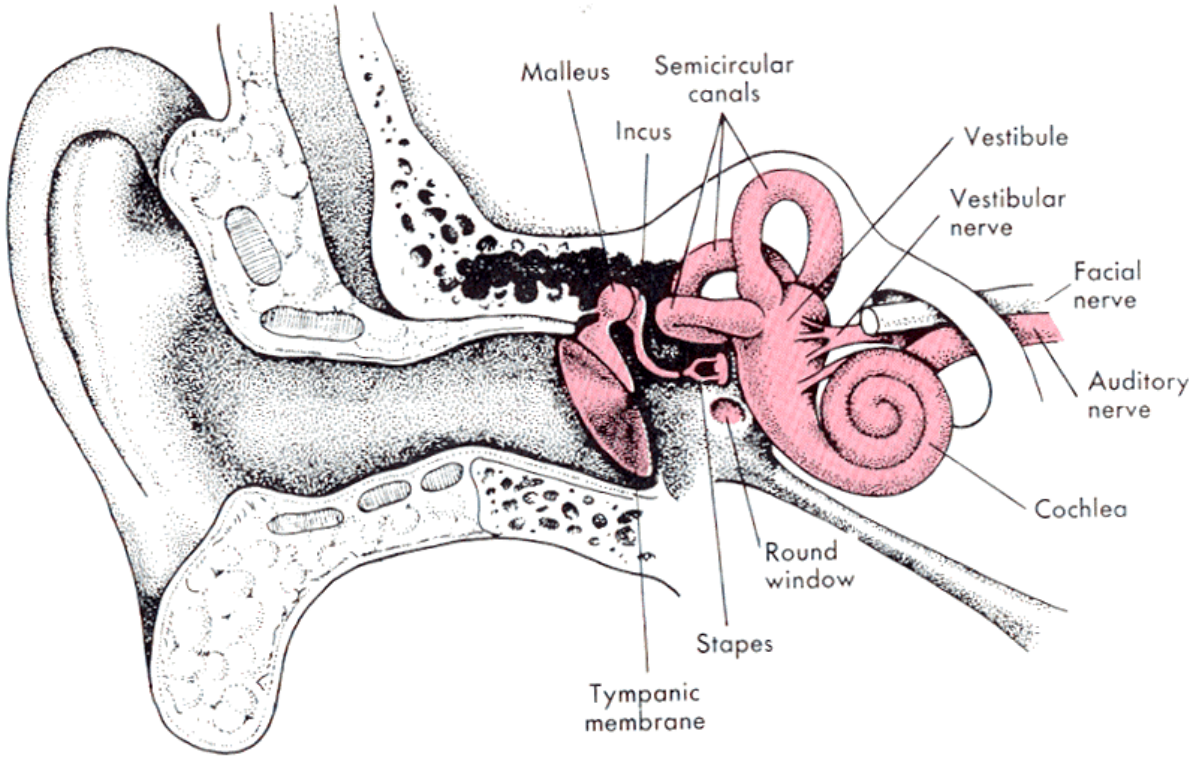


# Neural integrator




(Pastor, et al., 1994)

# Vestibular system




- Sense of balance and seasickness

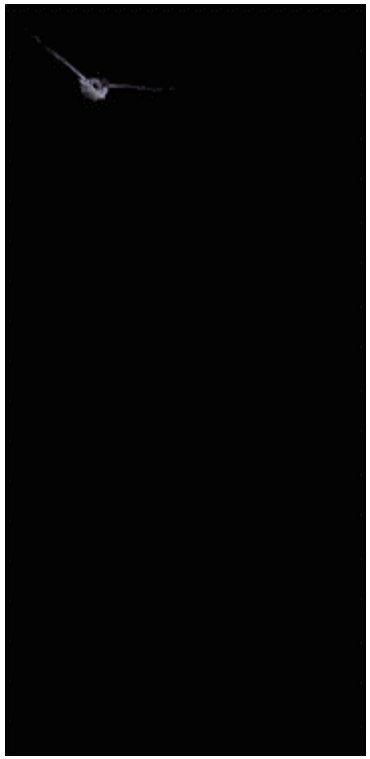
# Vestibular-ocular reflex



# Auditory localization

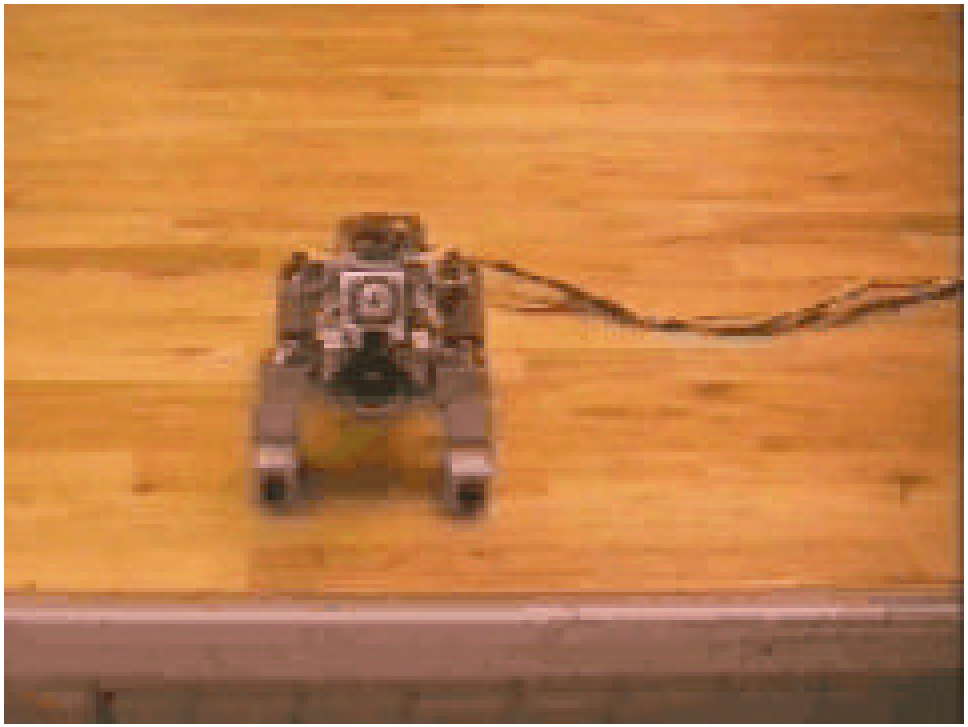


Barn Owl

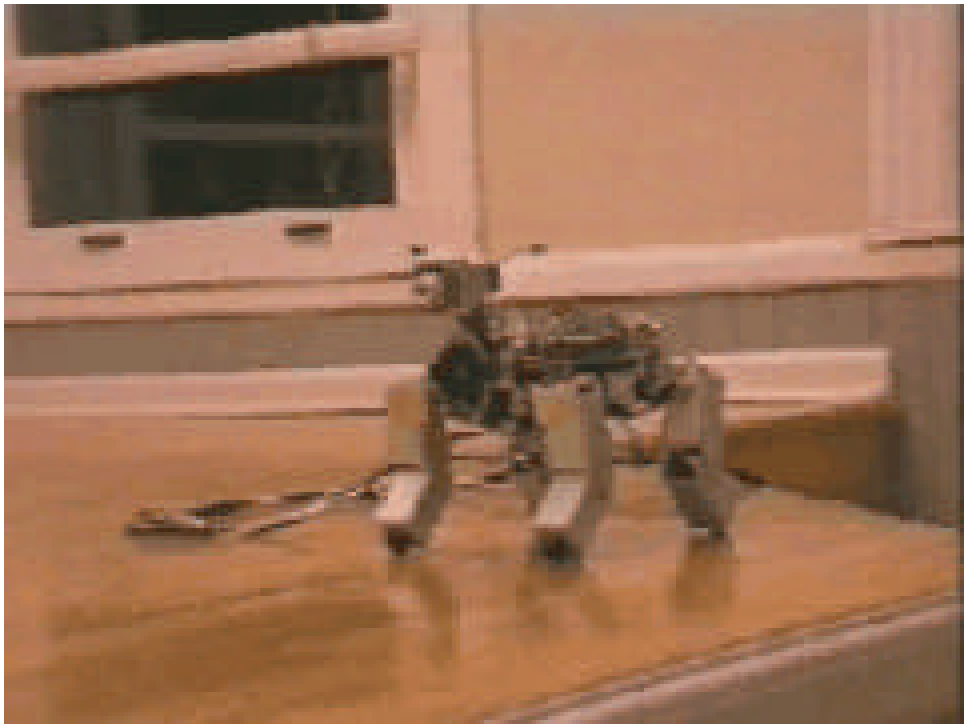


(Konishi, 1990)

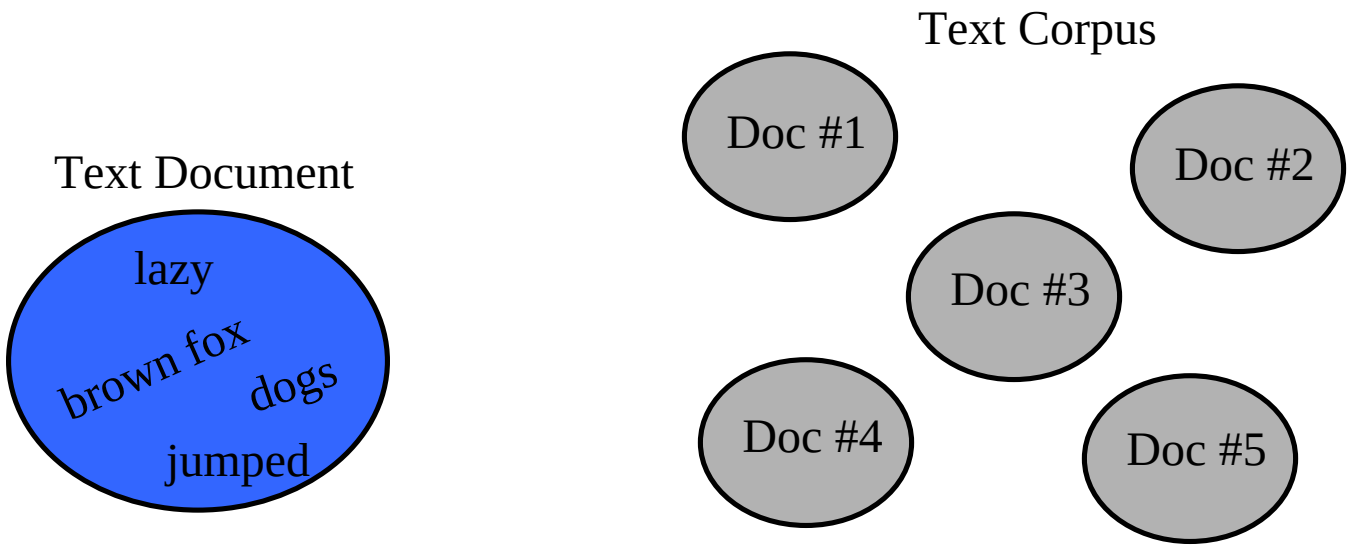
# Auditory localization



# Walking



# Language




- Model text document as collections of words.

# Text and images analogy

Word counts:

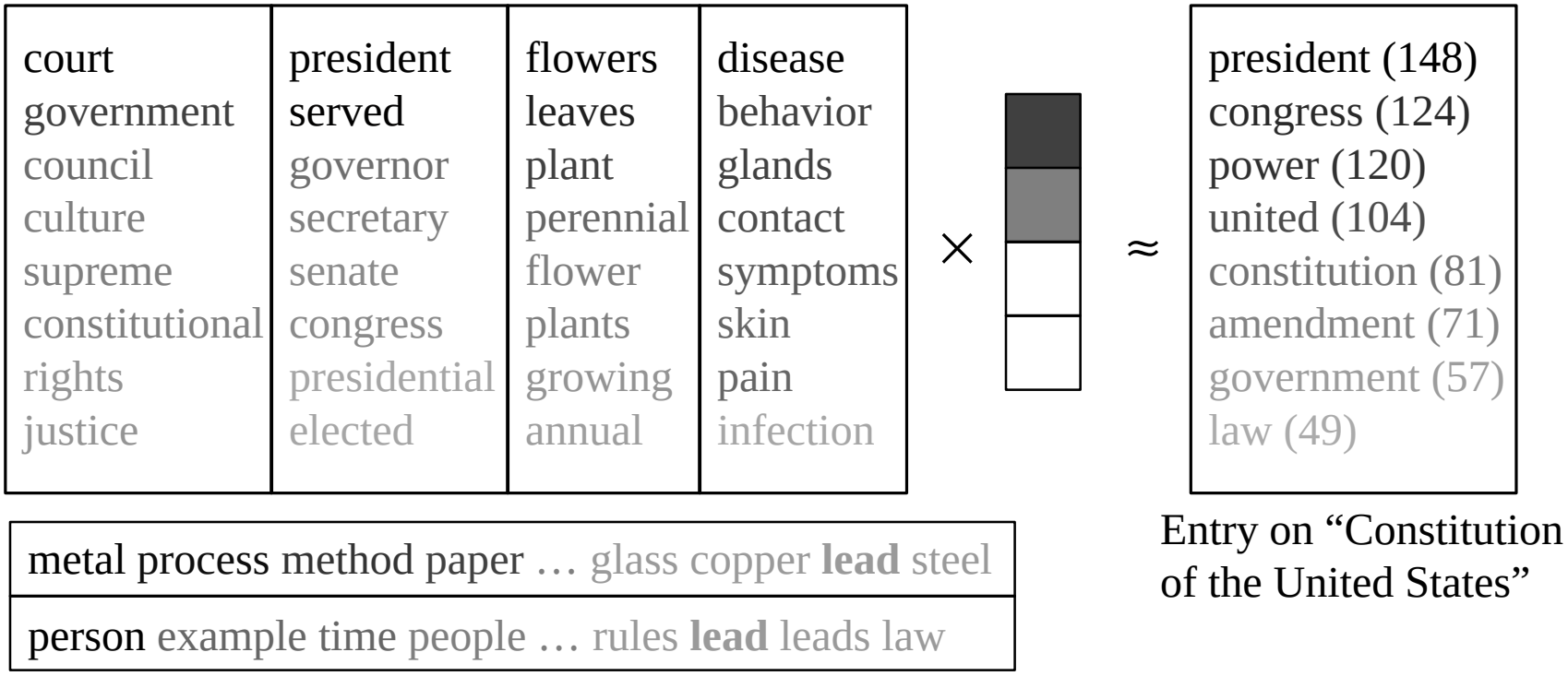
Documents

$$X = \begin{bmatrix} 1 & 0 & \star & 0 \\ 0 & 2 & & 1 \\ \star & & & \star \\ 1 & 0 & \star & 1 \end{bmatrix}$$ Words

| **Text** | | **Images** |
|---|---|---|
| words | ⟷ | pixels |
| document | ⟷ | picture |
| word frequency | ⟷ | grayscale intensity |

- Represent documents with word frequencies.
- Analogy between learning algorithms.

# Learned semantic topics




- Grolier encyclopedia: 15276 words, 30991 articles.
- Semantic features, word sense disambiguation.

# Multimodal integration



(Knudson, 1997)

- Vision, hearing and language combined

# Summary

- Adaptation and learning in biological systems important for vision, hearing, motor control.
- Mimic neural systems in computer algorithms.
- Robotic systems can learn from experience.
- But still cannot compete with your family dog or cat...