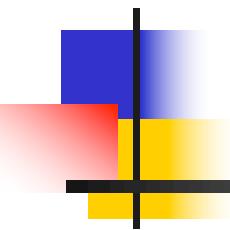




*Computational Perception*

&



# 生物信息处理

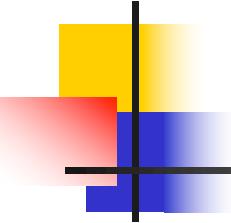
# **Computational Perception and**

# **Scene Analysis**

吴玺宏

北京大学智能学院

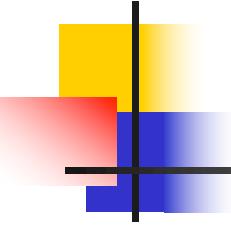
[wxh@cis.pku.edu.cn](mailto:wxh@cis.pku.edu.cn)



# 智能是什么？

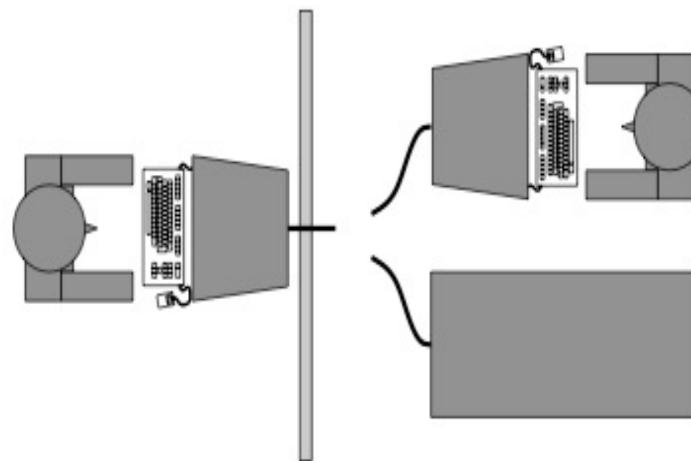
- 人工智能：由人类设计并在计算机上实现的模拟或再现某些智能行为的技术。

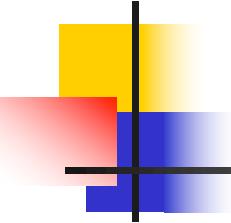
智能洗衣机是智能的吗？



# 智能是什么？

- Turing Test: Turing 提出利用语言的交流能力来评价智能水平。





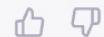
# ChatGPT测试——抽象问题

y

我信仰什么宗教好呢



信仰宗教是个人的选择，没有好或不好的说法。重要的是，宗教要满足个人的信仰需求，让他们有意义和安慰。每个人都应该自己思考自己是否需要信仰宗教，如果需要，选择适合自己的宗教。

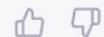


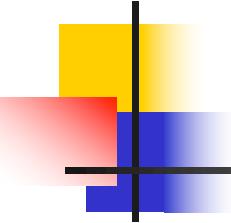
y

我不信仰宗教，能幸福吗



每个人对幸福的定义和追求都不相同，不信仰宗教也可以获得幸福。幸福来源于个人内心，它可以通过家庭、工作、朋友、兴趣爱好、志愿服务等方式获得。因此，信仰宗教或不信仰宗教并不影响一个人能否获得幸福。





# ChatGPT测试——具体动作

y

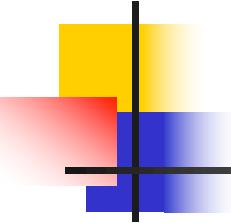
如何用筷子把花生米夹起来



用筷子夹花生米的方法一般如下:



1. 把筷子平放在桌面上，用另一只手拿起一些花生米，放到筷子上。
2. 将筷子斜着握紧，使花生米不被掉下。
3. 用另一只手的食指和中指将筷子上的花生米捏住。
4. 用嘴吃掉花生米。



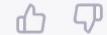
# ChatGPT测试——空间意象

y

我的鼻子和嘴的距离是多少



很抱歉，我无法告诉您您的鼻子和嘴的具体距离。因为人的身体特征是不同的，所以鼻子和嘴的距离也会有所差异。如果您需要知道您的鼻子和嘴的距离，可以向医生或身体测量专家询问。

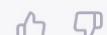


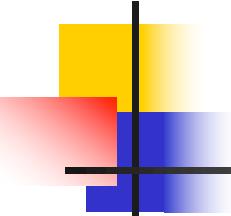
y

我的嘴是在鼻子的上方，还是下方



很抱歉，我无法告诉您您的嘴在鼻子的上方还是下方。因为人的身体特征是不同的，所以嘴的位置也会有所差异。如果您需要知道您的嘴在鼻子的上方还是下方，可以向医生或身体测量专家询问。





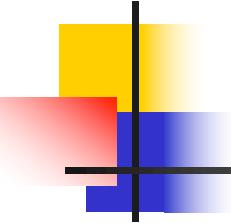
# 深度学习的缺陷

- 深度学习是上世纪80 年代兴起的技术

“灾难性遗忘”  
问题

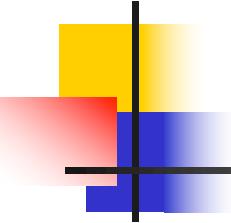
“推广性”  
问题

- “灾难性遗忘”问题：指在学习新知识时快速遗忘先前已学习知识的倾向
- “推广性”问题：依靠记忆，缺乏拓展性和鲁棒性。

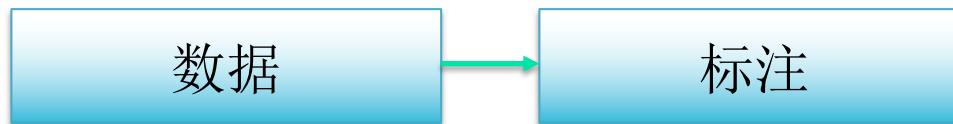


# 深度学习的缺陷

- 深度学习只是简单地相信所给的训练数据
  - 深度学习在不理解的情况下模仿数据中的内容，它不会否定任何数据，不会发现社会上的偏见，而只是“学习所有数据”。
  - 深度学习的公正性并非来自其自身，而是人类筛选和准备的数据。
  - 深度学习只相信它在数据中频繁看到的事物、底层模式和趋势，因此它会放大人类社会的偏见和问题。
- 没有一个深度学习网络（卷积、RNN、RNN + LSTM、GANs 等）可以解释其自身的决策。



# 深度学习陷入困境



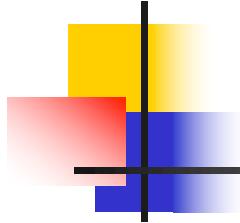
训练数据与实际应用的数据永远失配！是常态！

✓ 大数据时代，需要大规模标注数据；      ✓ **DNN**更深！更深！

“和面”方法

人工+智能！

危机：效能低于软件开发！

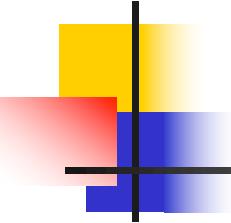


# 人工智能的几片乌云

符号落地

常识

主体：意图、情感



# 具身智能

- 研究机器人能够在**非结构**、**动态**、**部分观测**和**非确定**的环境下**自主**地生成目的并加以实现。



- **非结构**: 环境未知, 非特殊设计
- **动态**: 环境发生非预期的改变
- **部分观测**: 无法获得环境完整信息
- **非确定**: 传感器和动作执行存在噪声
- **自主**: 无人工干预情况下, 自己决策, 生成和完成任务

无人驾驶汽车

计算机程序

# AlphaGo vs. RoboCup



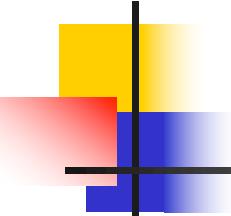
	环境	状态	状态改变	获取信息	传感器信息	决策方式
围棋	结构化	离散	回合	完全	确定符号	集中总控
RoboCup	非结构化	连续	实时	不完全	非确定信号	分布自主

# 飞机 vs. 飞鸟



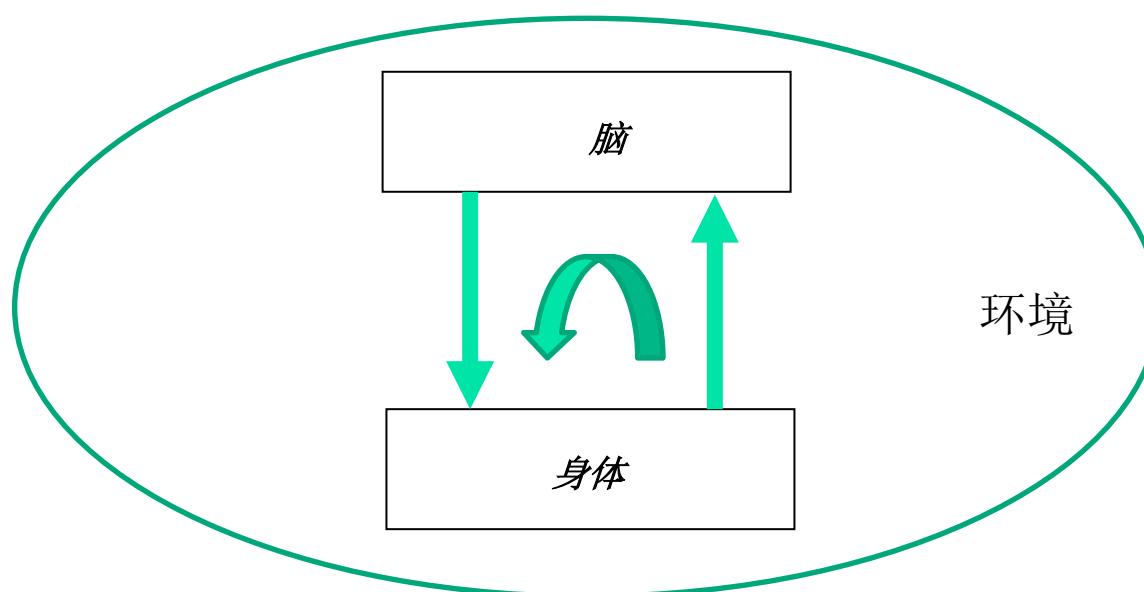
- 感知-运动的协同控制、进化
- 感知信号与先验知识
- 环境与行为的适应
- 感性-理性存在鸿沟
- 目的和动机：自主行为

具身=>大脑=>环境复杂性+价值

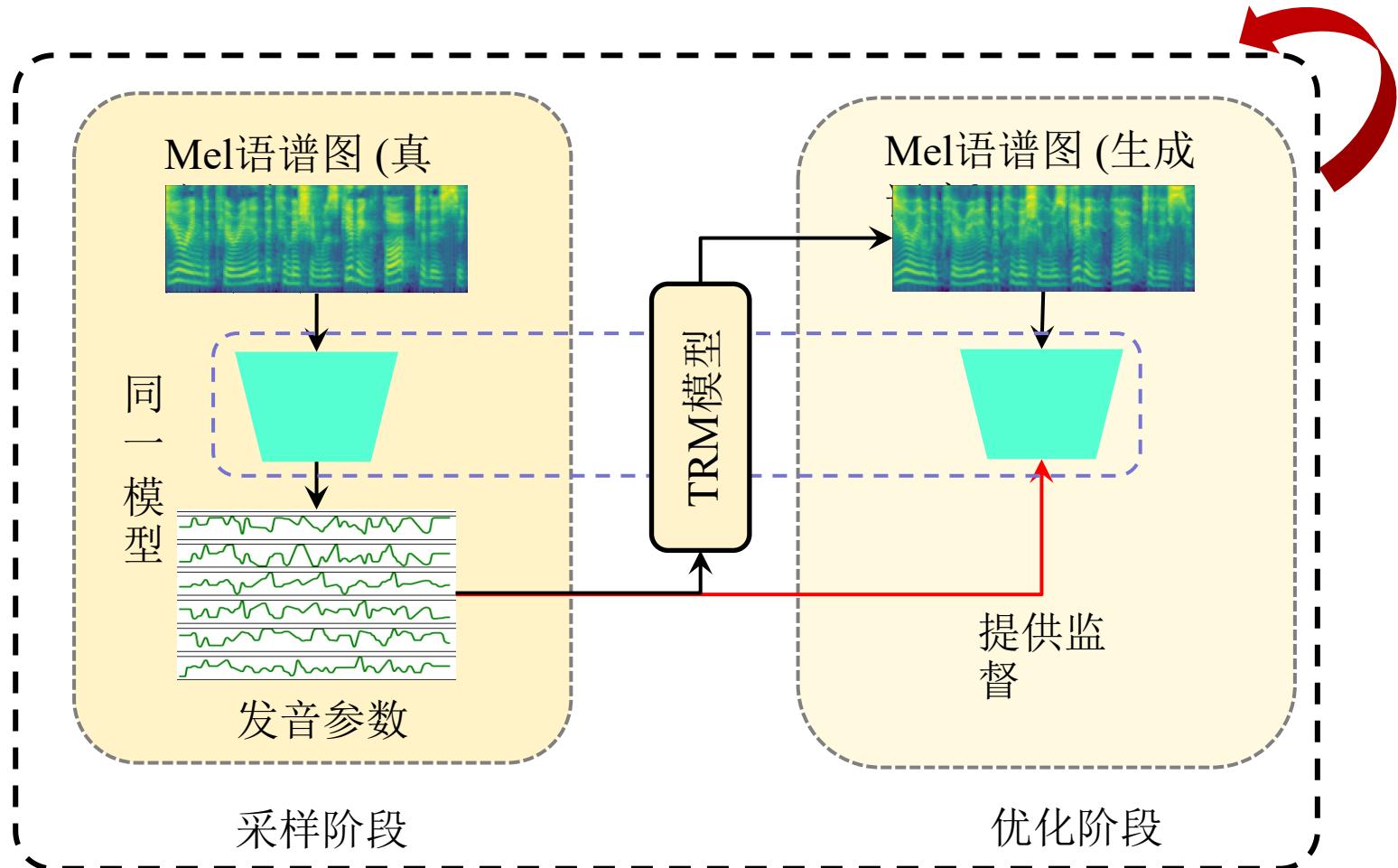


# 具身智能+类脑智能

- 通过逆向推断和正向生成的协同，不断调整脑内部网络，以实现内部生成世界与外部客观世界的一致。



# 具身自监督学习框架



# 演示：呼麦及其产生机制

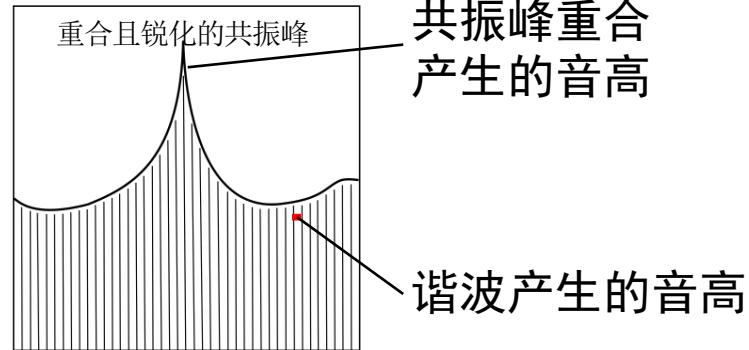
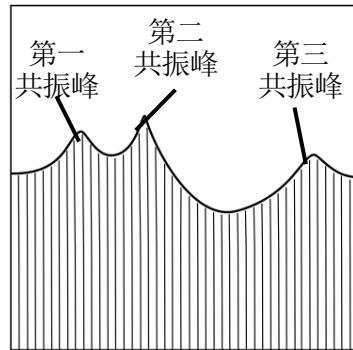
## ■ 呼麦

- 蒙古族、图瓦人的一种民族音乐形式
- 单个歌者同时产生两个音高



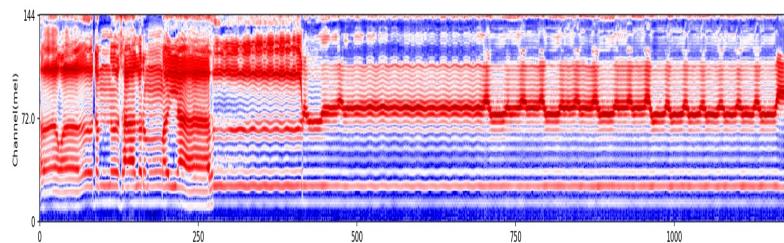
## ■ 产生机制：

- 一种运用泛音的歌唱方式，第一、第二共振峰重合

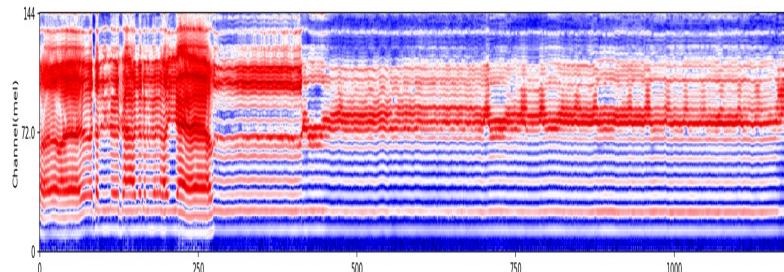


问题：发音器官采取何种姿态可使得前两个共振峰重合？

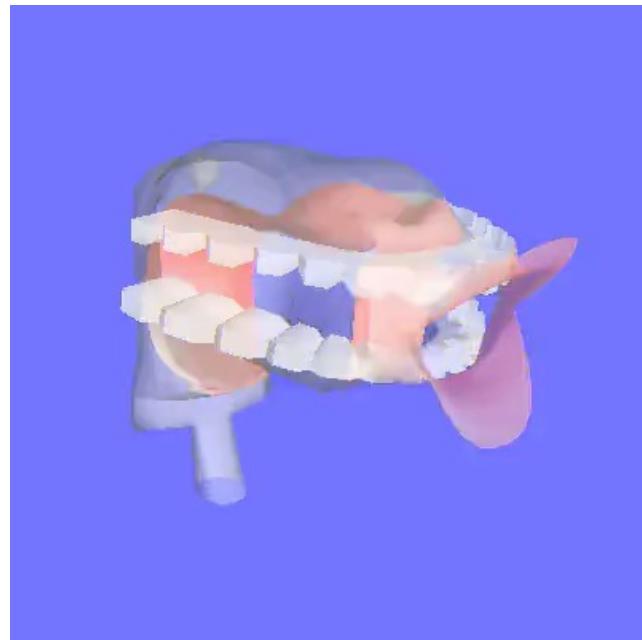
# 演示：呼麦反演



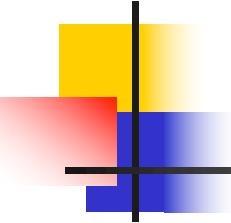
(a) 实录呼麦的部分语谱



(b) 重构呼麦的部分语谱

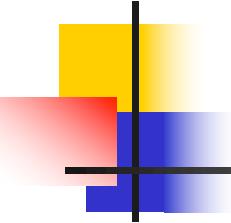


重构的呼麦与估计的发音动作

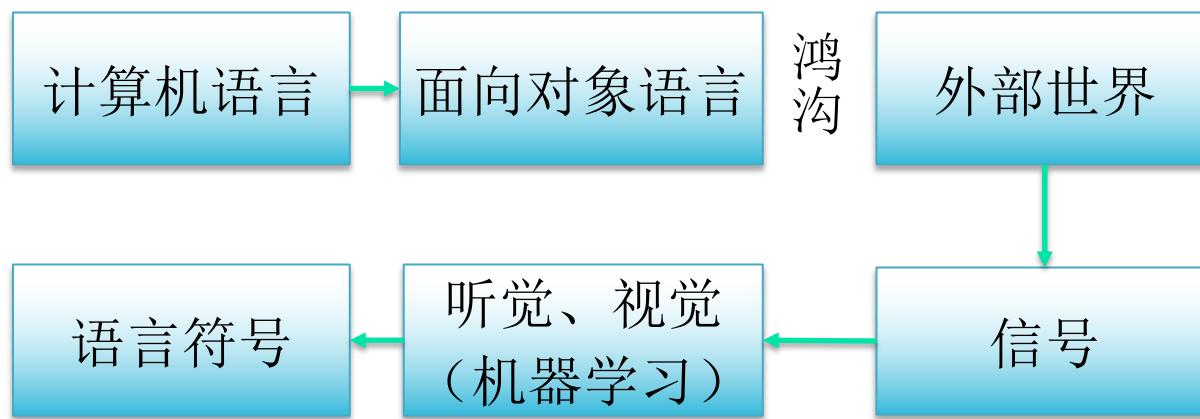


# 智能学科的定位——横断科学

- 数学的微积分方法、物理学的建模：研究世界（物质、事务）的运动和变化规律；
- 计算机：基于获得数据，实现离线的运算过程、业务流程；
  - 图灵机：程序逻辑（预制的条件控制、循环控制）；
- 自动化：针对确定环境、特定任务，实现在线的反馈控制，进而预测
  - 维纳控制论；线性系统；单任务、缺少主体；
- 智能：不仅能适应动态环境，还可产生新的动机和目的，进而改造环境；并通过语言交际实现知识的传递。
  - 机器学习理论=>具身智能+类脑智能+语言智能

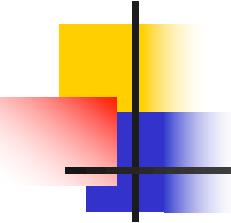


# 计算机 vs. 智能

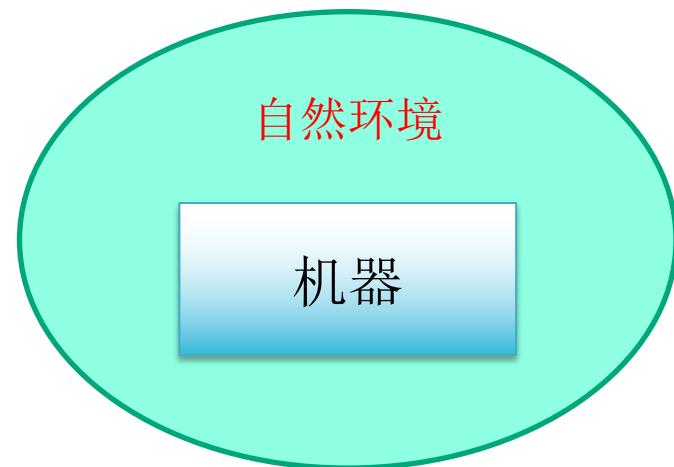
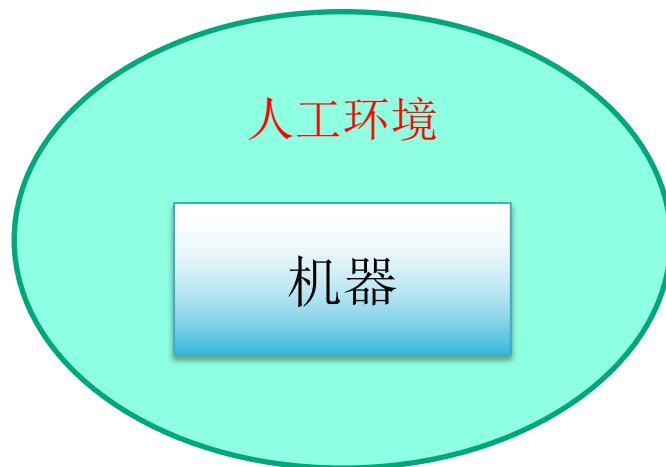


符号? 指向性问题





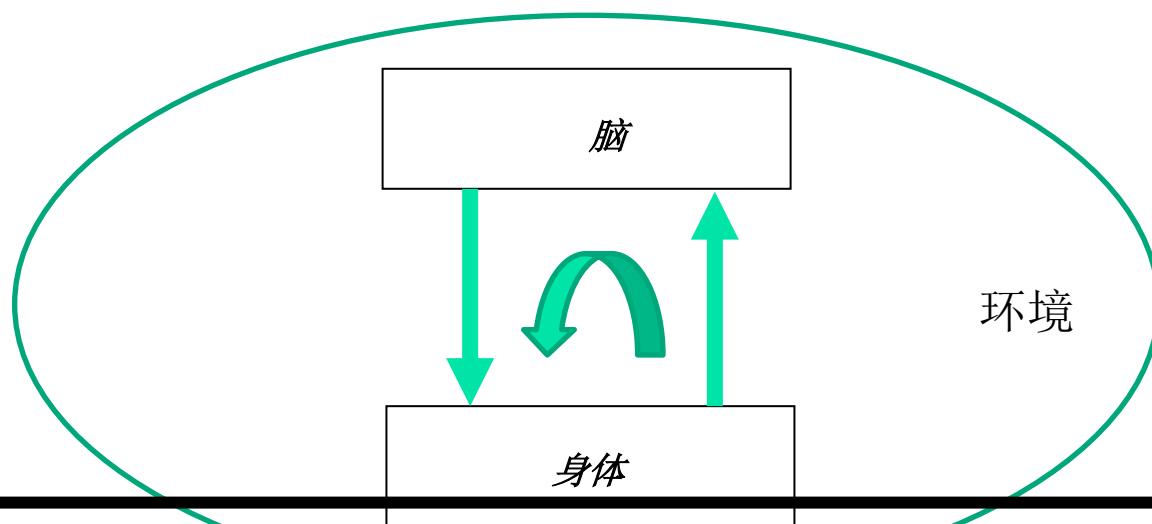
# 自动化 vs. 智能



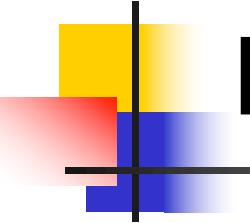
自然环境：复杂、多变、未知

# 具身智能+类脑智能

- 通过逆向推断和正向生成的协同，不断调整脑内部网络，以实现内部生成世界与外部客观世界的一致。



复杂环境中的知觉问题



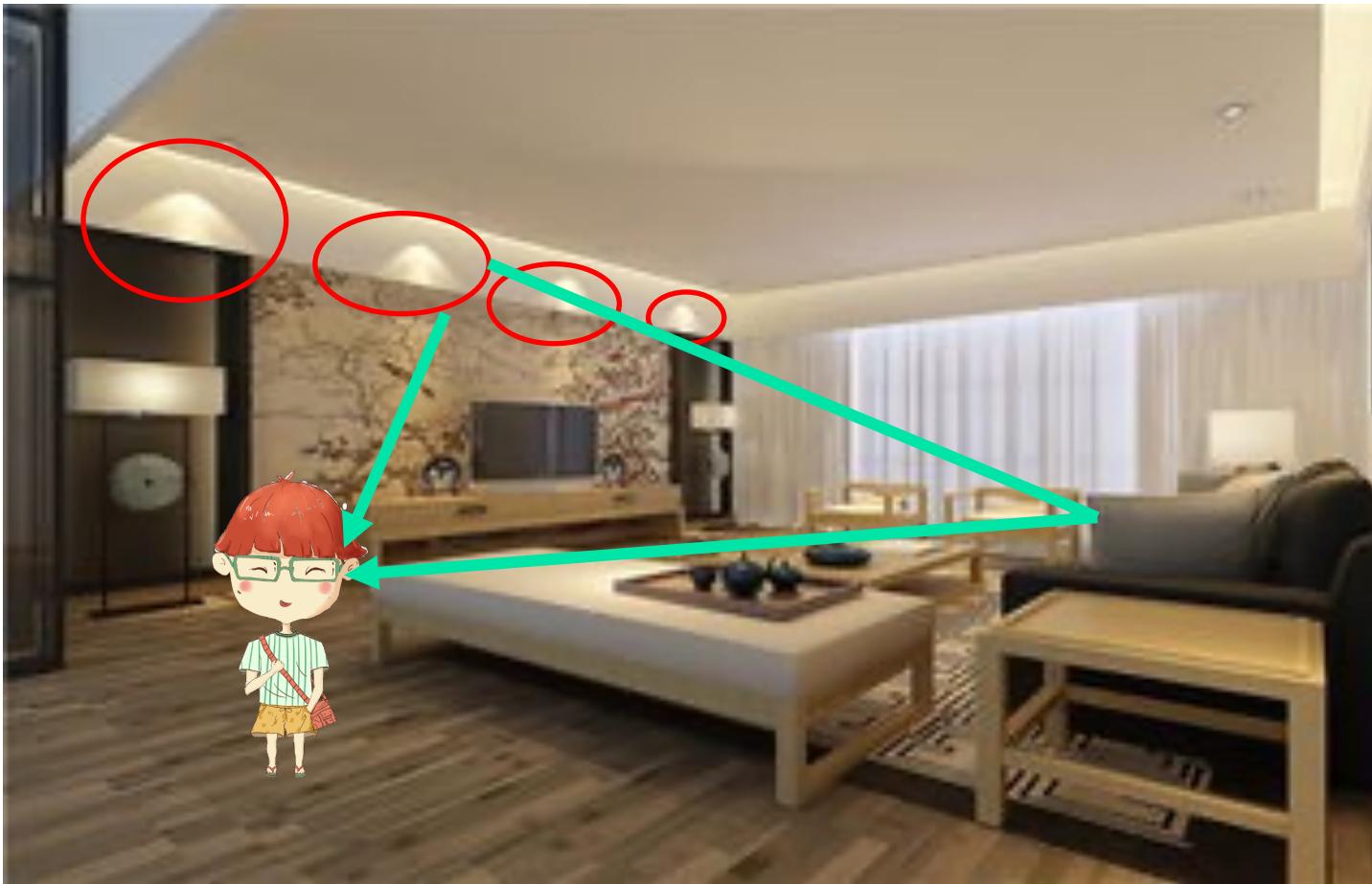
# Perception

---

- Perception:

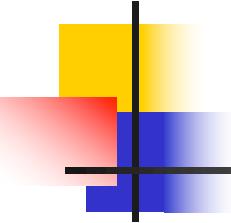
The process of **acquiring knowledge** about **environmental objects and events** by **extracting information** from the light or sound they **emit or reflect**.

# information from the light



# information from the sound

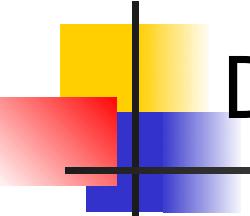




# Natural Perceptual Behavior



from Rodieck, 1998



# Decomposing Natural Perceptual Behavior



**Sensation**



**Orientation**



**Perception**

from Regan, 2000



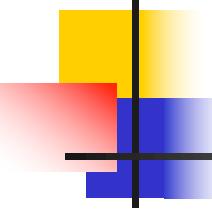
**Search**



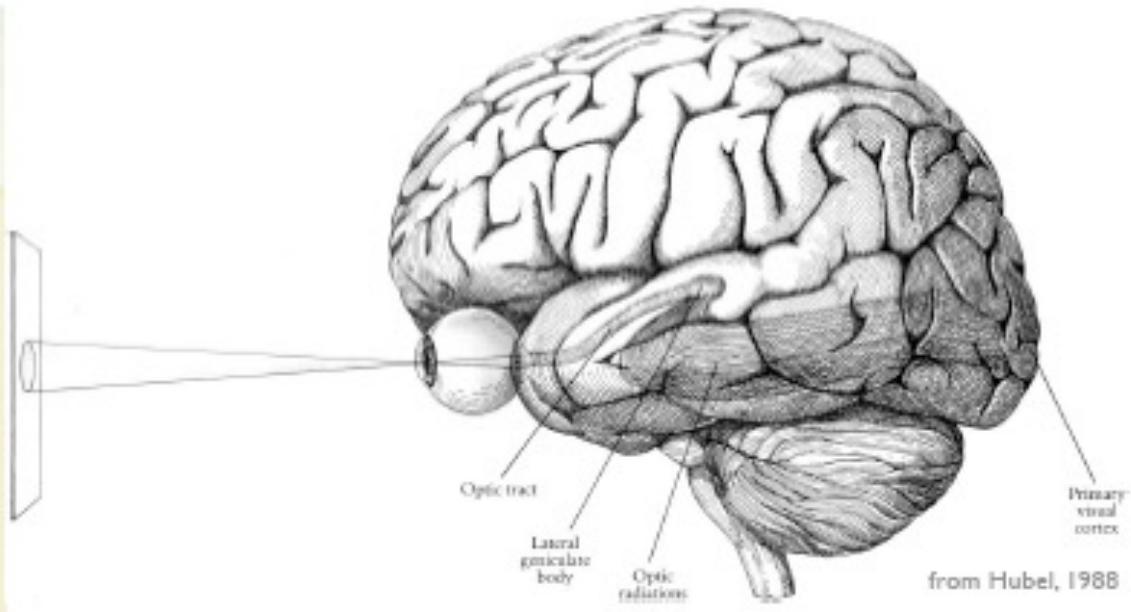
**Separation**



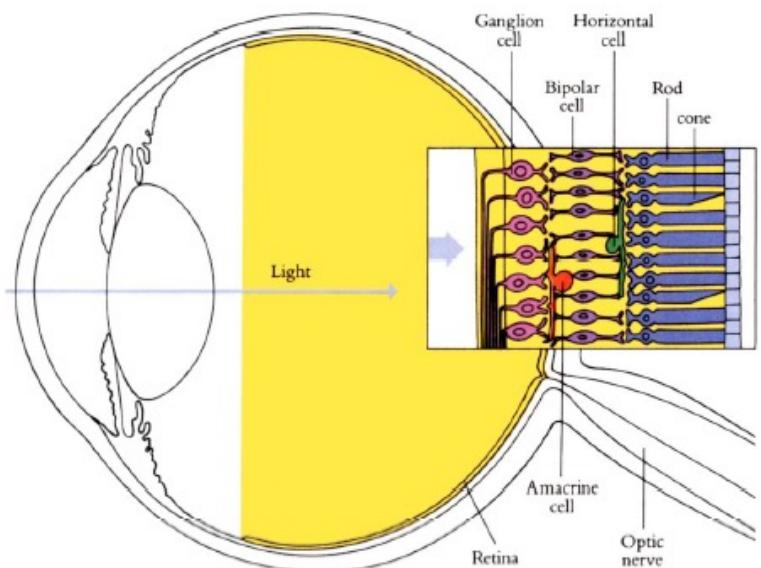
**Identification**



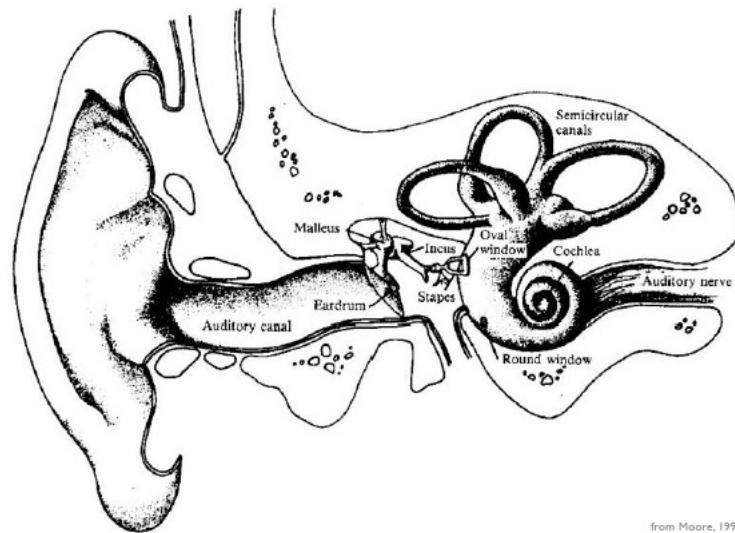
# What are the computational principles of perceptual systems?



# It starts here



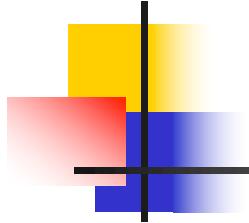
from Hubel, 1988



from Moore, 1997

Visual coding

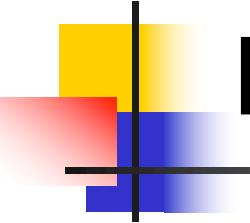
Auditory coding



# How general are features?

---

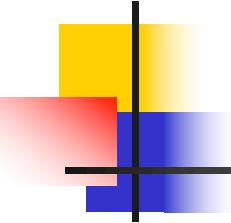




# How general are features?

---

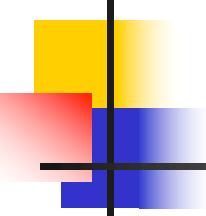




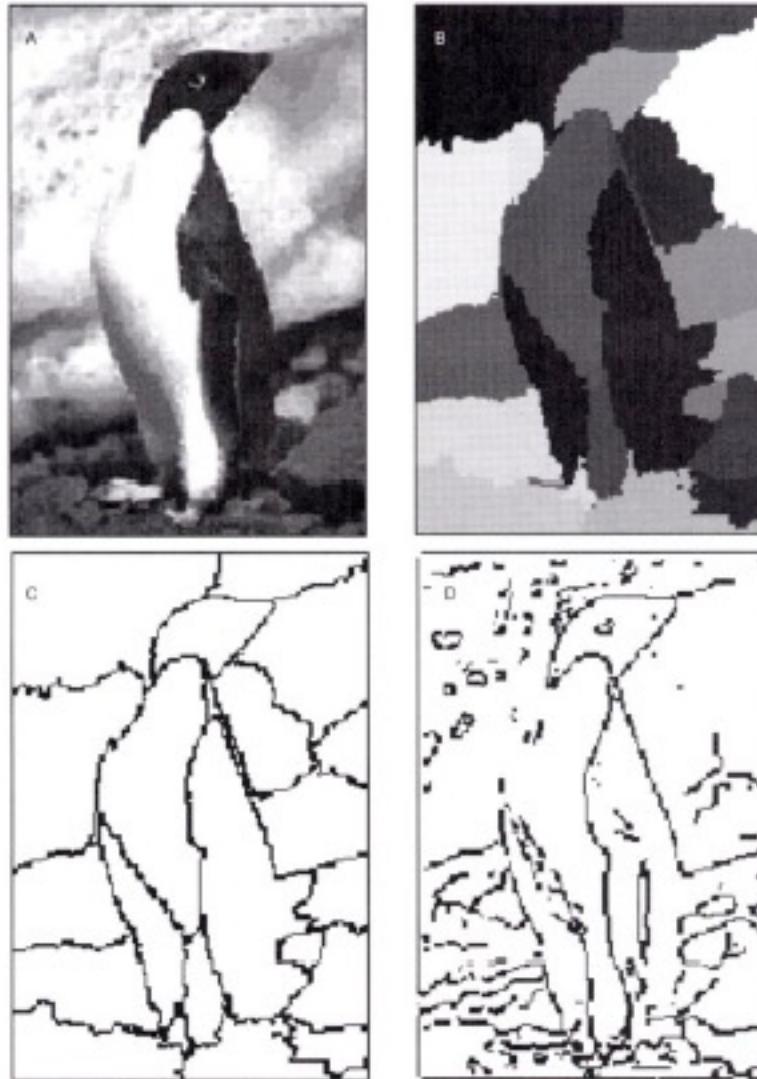
# Grouping parts: more difficult than it appears



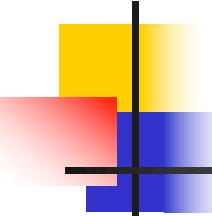
How should “parts” be grouped?



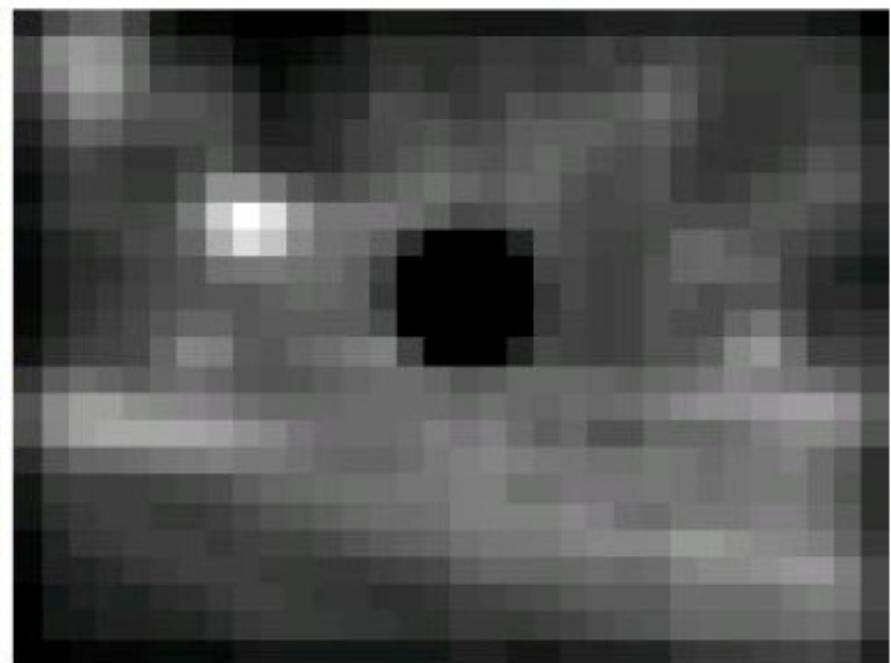
# Segmentation algorithms often fail

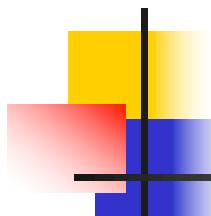


from Leung and Malik, 1998

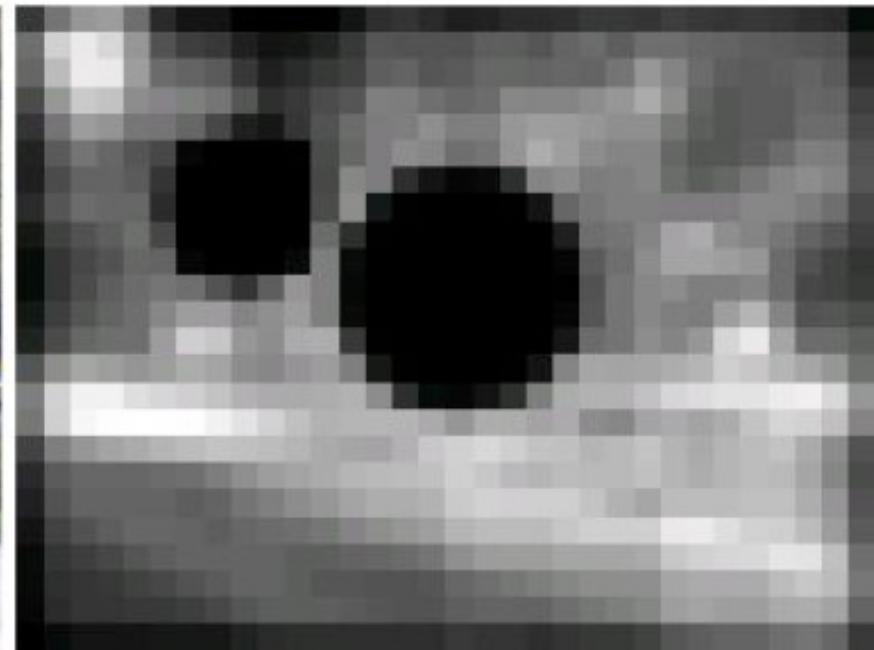


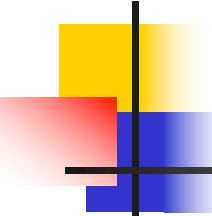
# Selecting information through attention



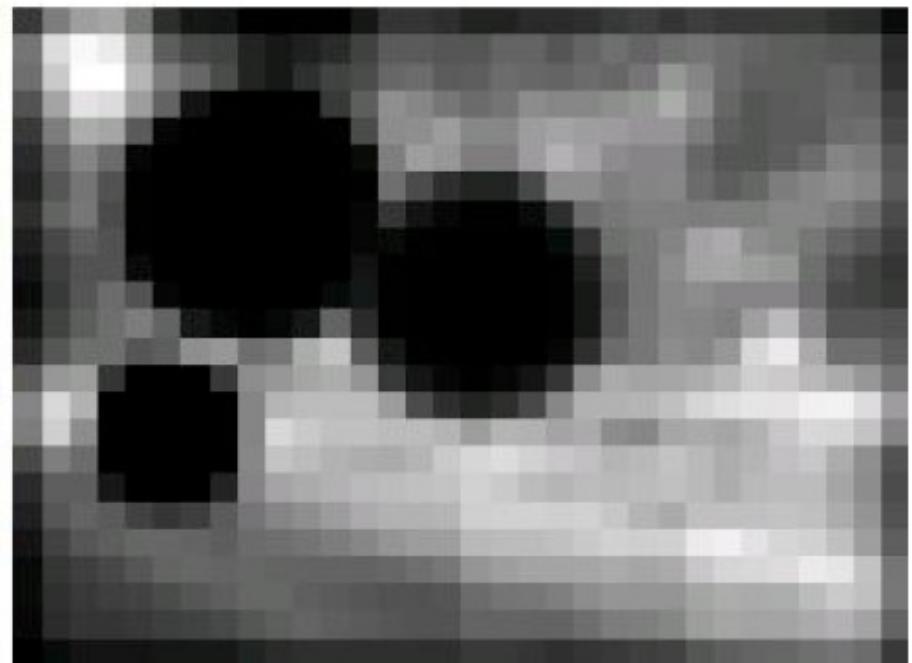
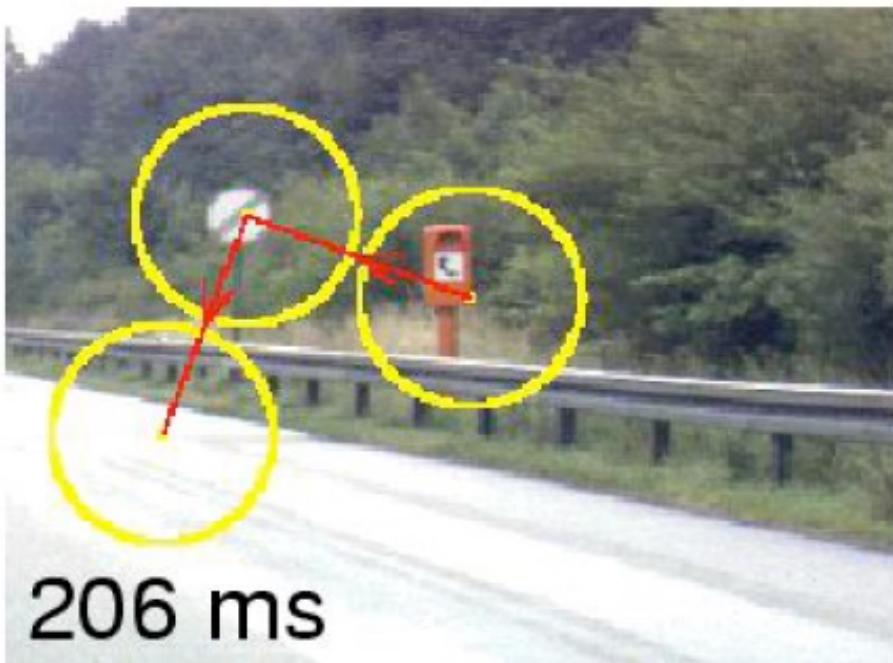


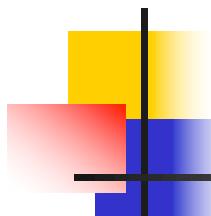
# Selecting information through attention



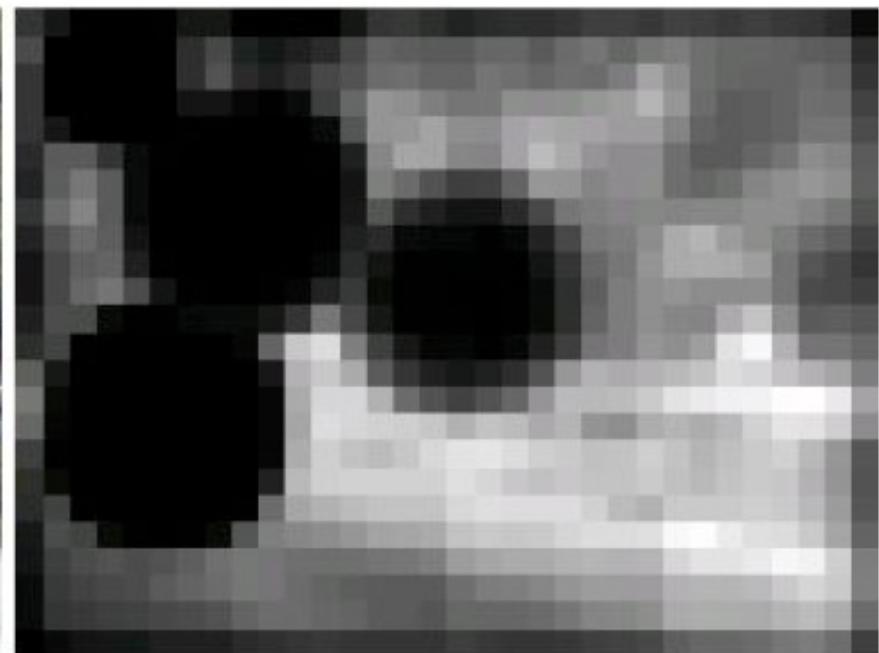
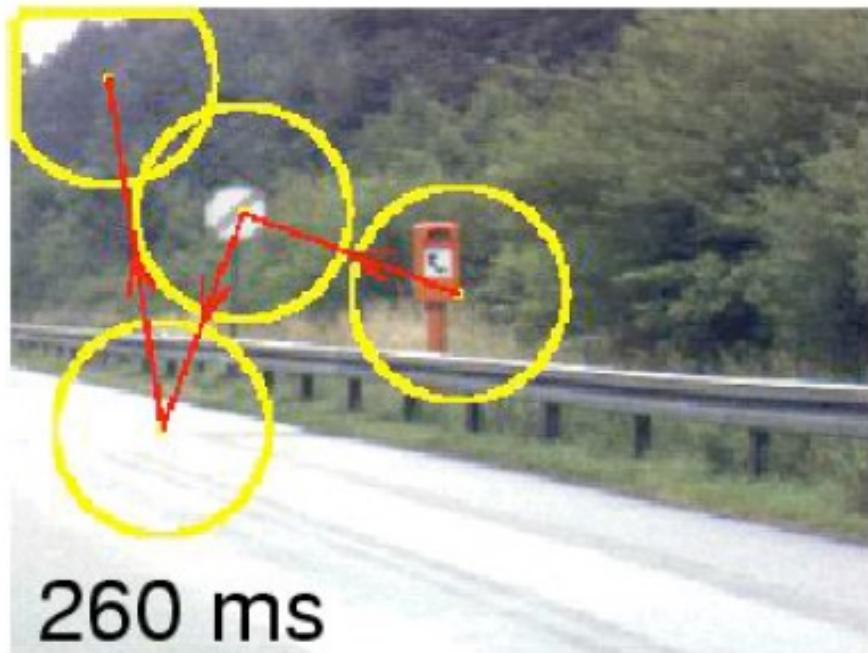


# Selecting information through attention

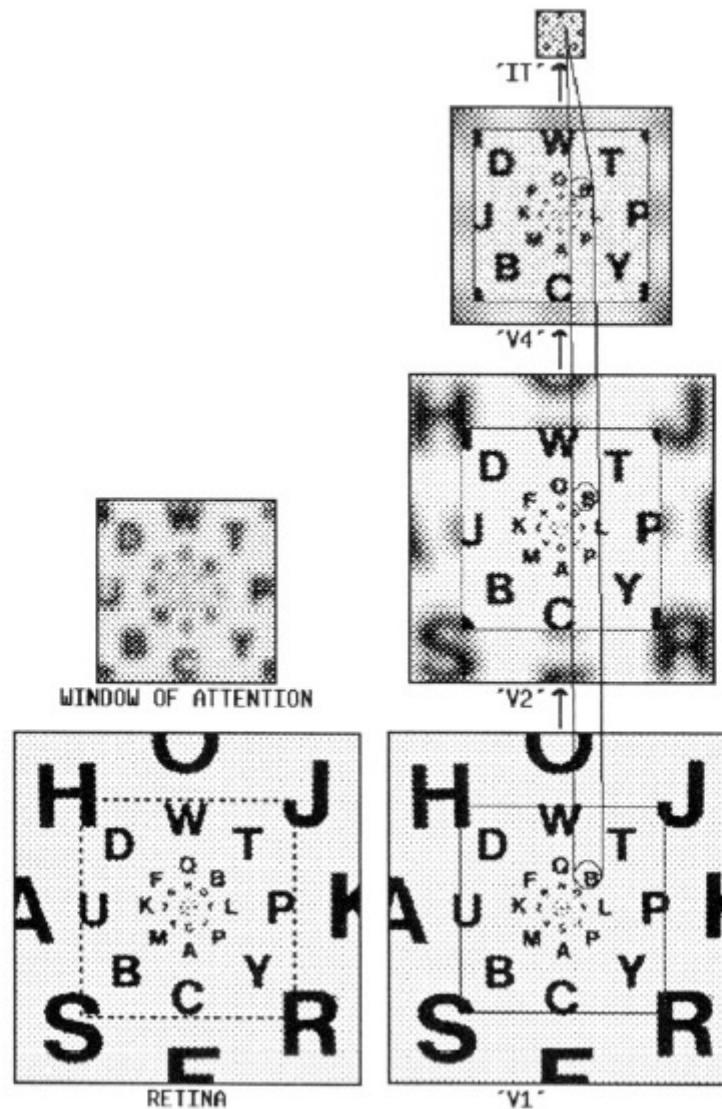




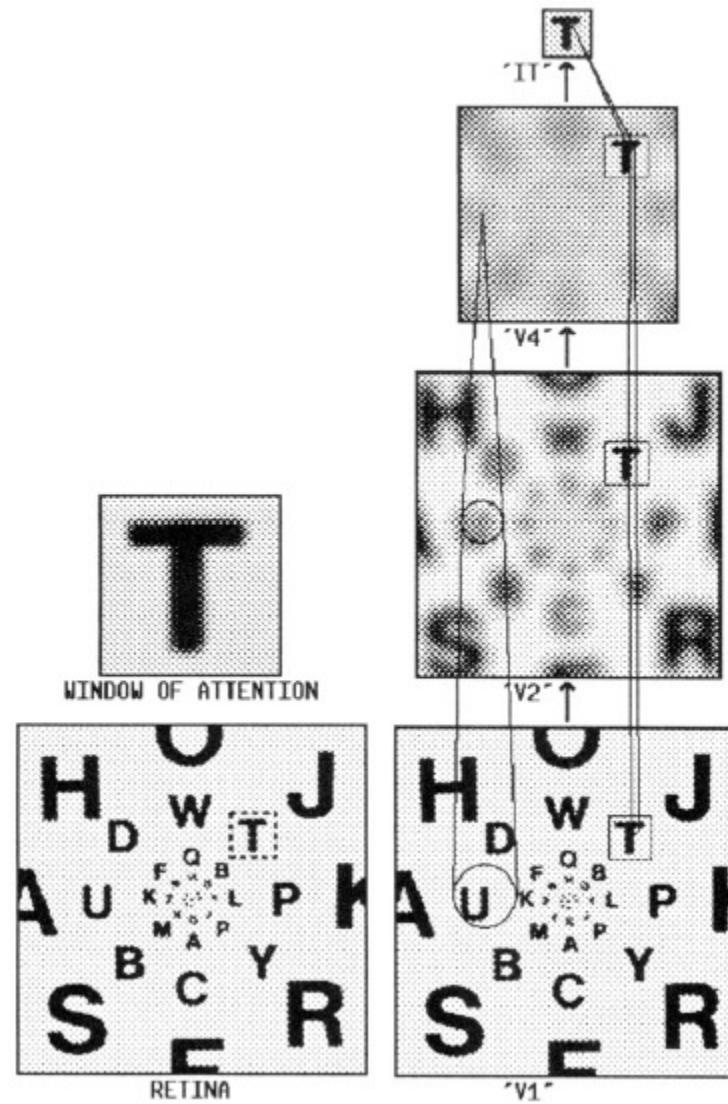
# Selecting information through attention

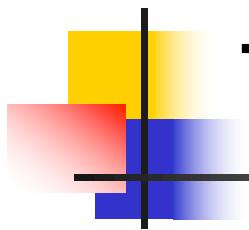


# Selecting information through attention



# Selecting information through attention

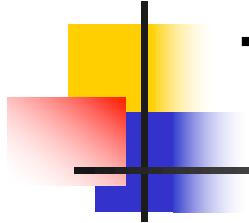




# The importance of context

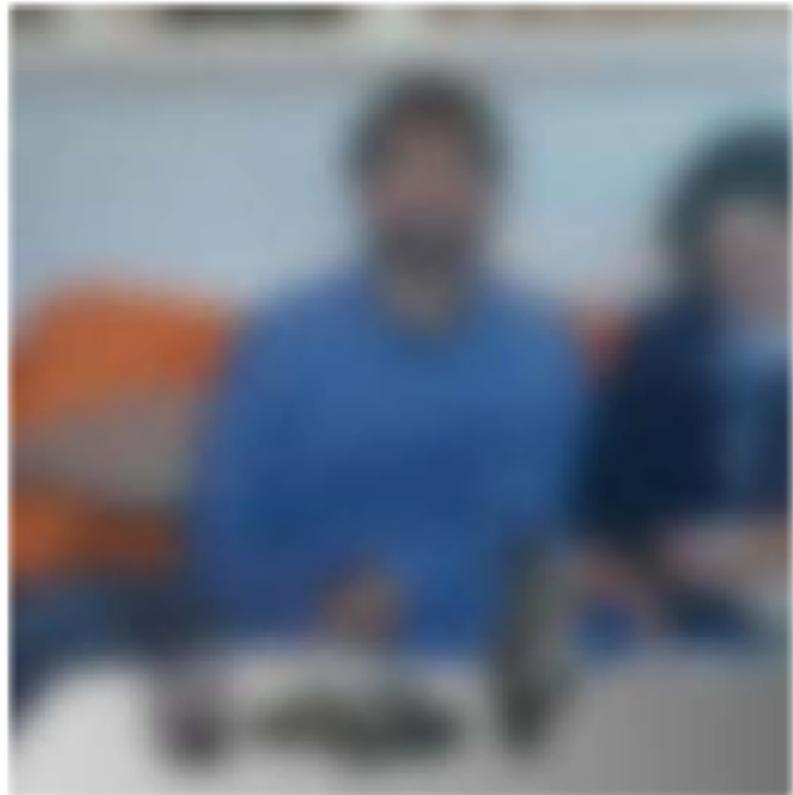
---

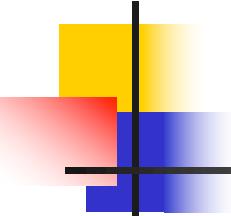




# The importance of context

---

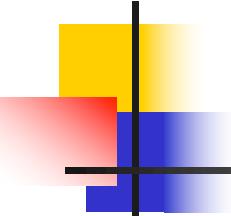




# Object recognition



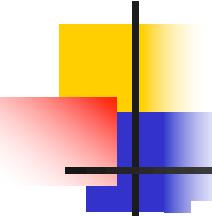
from Lowe, 2003



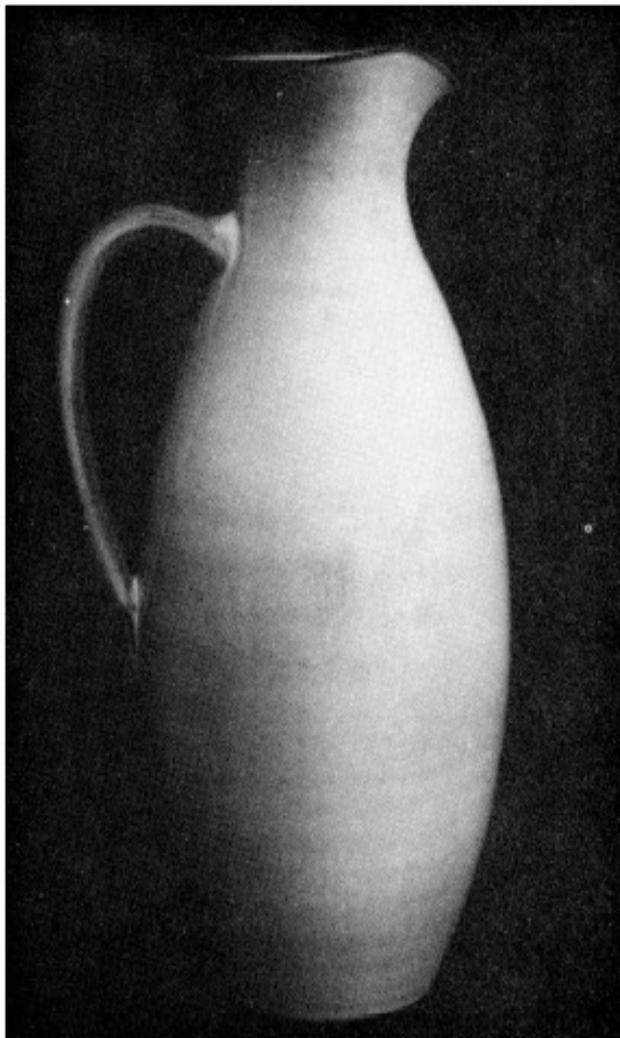
# Object recognition by features



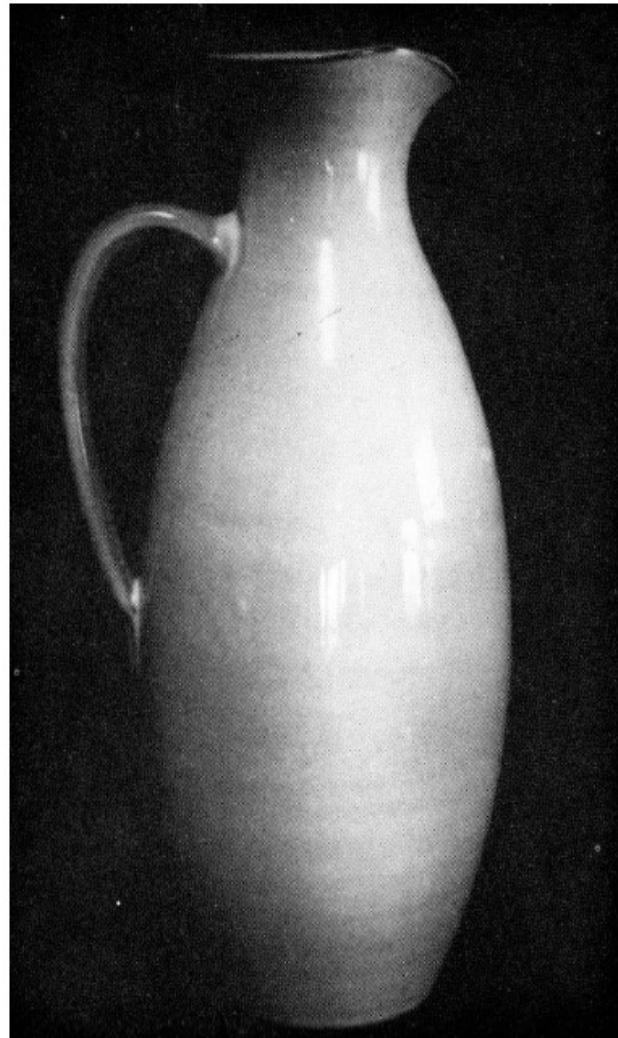
from Lowe, 2003



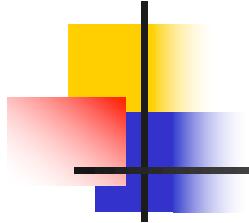
# Non-locality in visual perception



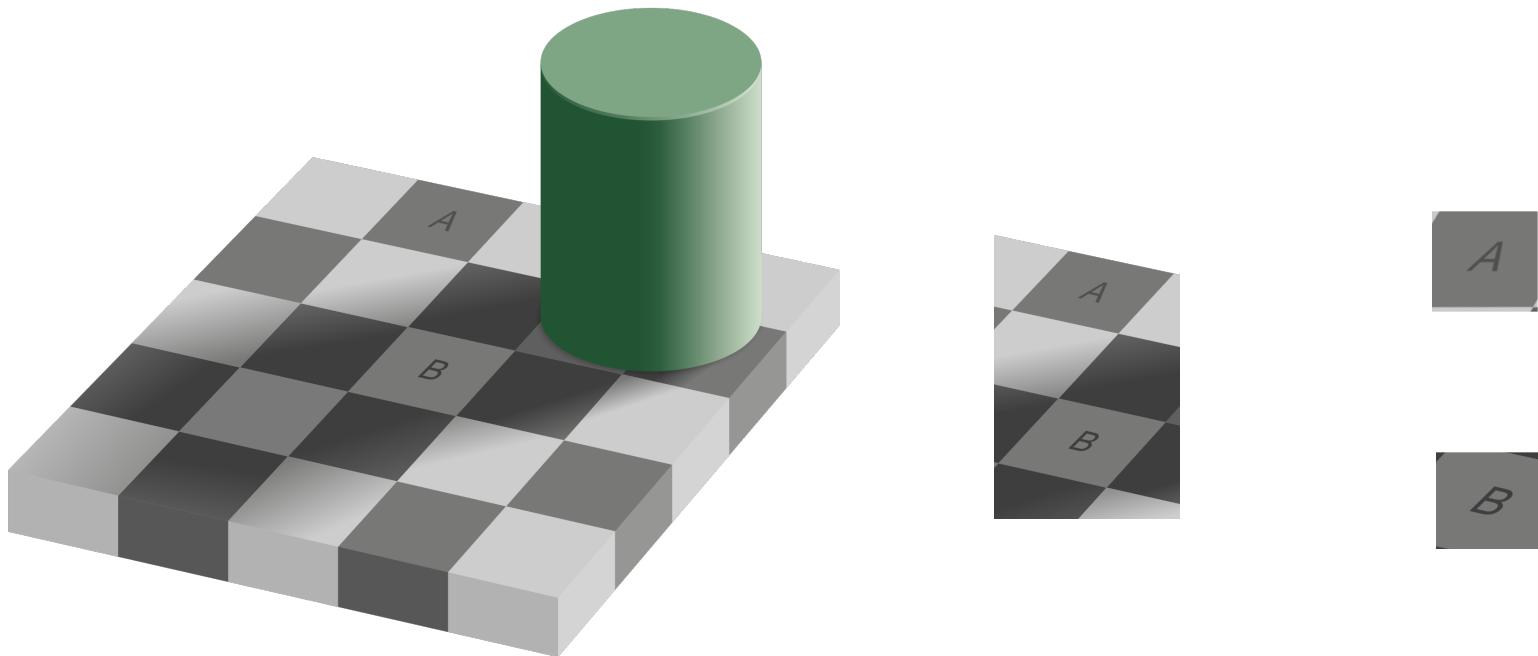
from Gold:

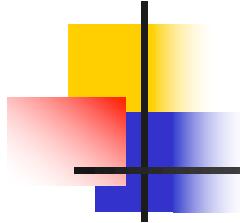


from Goldstein, 2001



# Non-locality in visual perception



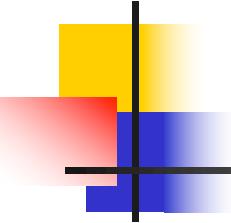


# Non-locality in visual perception

---



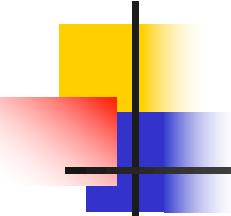
from Rock, 1995



# Methodology

---

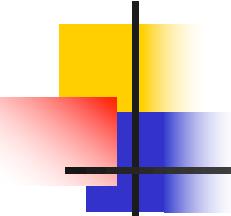
- Science: ethology, psychology, psychophysics
- Theory: modeling, idealization
- Engineering: computation, algorithms
  - Digital Image Processing
  - Speech Signal Processing



# Theory, Model and Algorithm

---

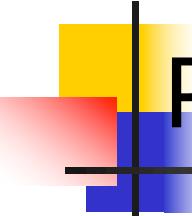
- A **theory** is an integrated set of statements (called hypotheses) about underlying mechanisms or principles that not only organizes and explains known facts, but also makes predictions about new ones.
- Solving the right problem
  - Appropriate level of abstraction
  - Model scope and limitations
  - Relevant algorithms



# The evolutionary utility of perception

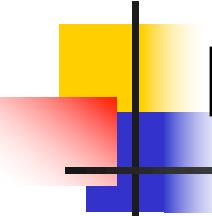
- Perception evolved to aid in the survival and successful reproduction of organisms.
  - Navigation of bats
  - Veridical perception
- Evolutionary purpose of perception:
  - To make contact between the needs of the organism and the corresponding opportunities available in its environment.
  - Human – computer – robot

推荐：达马西奥， 《当自我来敲门：构建意识大脑》



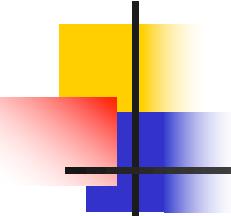
# Perception as modeling the environment

- The observer is constructing a **model** of what environmental situation might have produced the observed pattern of sensory stimulation.
- The perceptual models must
  - Be closely coupled to the information in the projected image of the world
  - Provide reasonably accurate interpretations of this information.
- Modeling:
  - Visual completion: self-occluded surfaces
  - Predicting the future: to compute dynamic, predictive models of the environment.



# Perception as apprehension of meaning

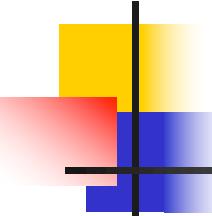
- We perceive an object not just as having a particular shape and being in a particular location, but a person, a dog, a house...
- Being able to classify (recognize or identify) objects as members of known categories allows us to respond to them in appropriate ways.
  - Classification: foreign object
  - Attention and consciousness
    - Stimulus-driven process
    - Cognitive constraints: higher-level goals, plans and expectations



# Inverse Problem

---

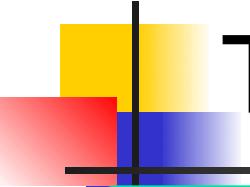
- How to get from optical images of scenes back to knowledge of the objects that give rise to them?
- The mathematical relation between the environment and its projective images is not symmetrical.
  - Well-defined – underspecified
- Perception is a heuristic process.
  - Inferences are made about the most likely environmental condition that could have produced a given image or sound.



# Inverse Problem

---

- That is why auditory and visual perception is so complex.
- Were it not for the fact that our brains manage to come up with the correct solution most of the time, it would be tempting to conclude that 3-D auditory and visual perception is simply impossible!

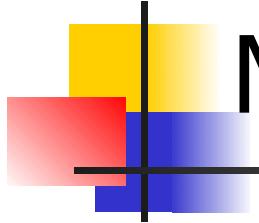


# Topics

- Overview and general issues
- Theoretical approaches to perception

- Color vision
- Image-based spatial processing
- Perceiving surfaces oriented in depth
- Organizing objects and scenes
- Object perception
- Perceiving motion and events

- Auditory representation
- Pitch and timbre
- Binaural information processing and sound source localization
- speech perception and language acquisition



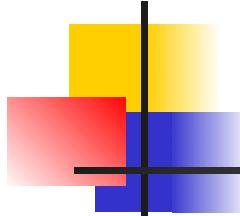
# Mechanism and Computation

---

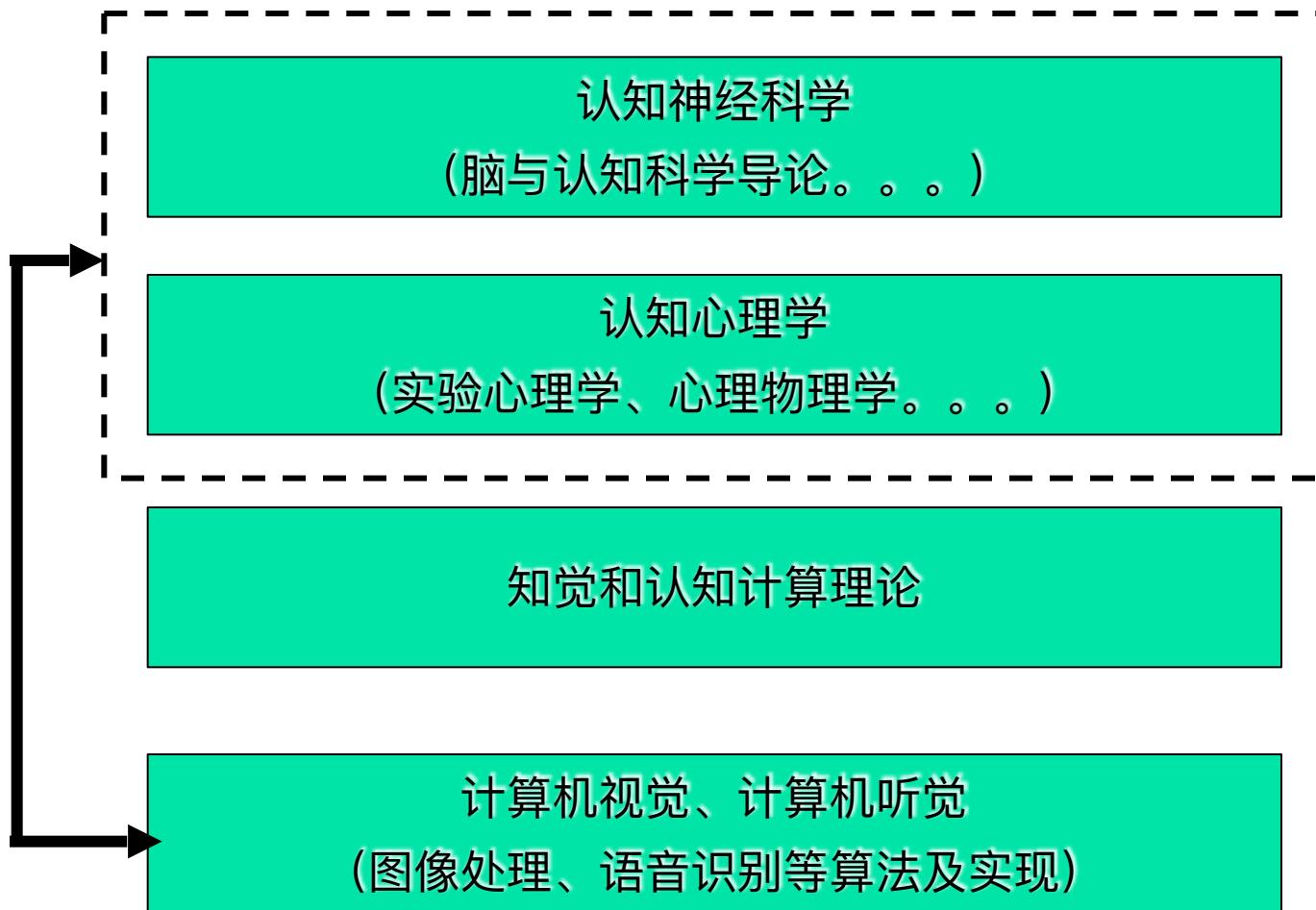
Physiological Mechanisms

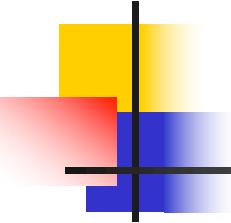
Psychophysical Mechanisms

Computational Approaches



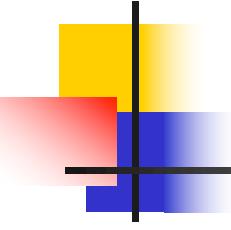
# Relationship with other courses





# References

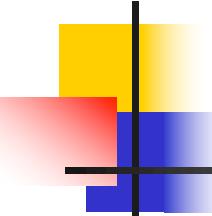
- Y.Ando, P.Cariani, Auditory and visual sensations, Springer
- S.E. Palmer, Vision science – Photons to phenomenology, MIT press
- A.S.Bregman, Auditory scene analysis – The perceptual organization of sound



# 考核

---

- 出席及表现: 15%
- 编程作业: 25%
- 考试
  - 期中: 25%
  - 期末: 35%



# Conclusion

---

- Natural perceptual behaviors
  - Sensing, orienting, perceiving, searching, separating, identifying
- The computational problems
  - Feature, part-grouping, selective attention, scene context, local-global
  - evolutionary utility, constructive act, modeling the environment, apprehension of meaning
  - Inverse Problem
- Course topics