# Computational Perception
# 15-485/785

# Class topics, readings, and references

## – *tentative* –

This detailed syllabus provides a list of the topics we will cover in the class. Each section includes a list of handouts for that topic, which will include reading assignments and background materials. The background handouts are not required reading, but are provided for you reference in case you haven't covered the material in another class. These handouts are drawn from a wide range of sources. Some will provide background for the material covered in lectures, others will provide the background for the homework assignments. The readings for each topic are required and will be discussed in class. The supplemental readings are provided if you wish to pursue certain topics in greater depth.

The topics are organized roughly as the flow of information in our own perceptual systems, i.e. starting with the sensory information and moving on to increasing levels of abstraction and complexity. Although we start with sensory coding and end with object recognition, throughout the course we will be discussing all levels of perceptual processing and computations, because as you will learn, they are highly interrelated. This course will also cover both audition and vision, which necessitates some jumping back and forth, but they are more related than you might think, because there is considerable overlap in terms of the fundamental problems solved in each modality.

# 1 Course Overview

## 1.1 Topics

The first lecture provides an overview of the course and discusses its goals, approach, and organization.

## 1.2 Handouts

Course syllabus.

## 1.3 Readings

Nakayama gives a sweeping and insightful historical overview of the study of visual perception. He discusses the rise (and fall) of many of the most important scientific approaches and schools of thought, citing their key contributions to scientific development of the field as well as their ultimate conceptual limitations.

von Békésy's introductory chapter to his book gives an excellent an entertaining discussion of scientific philosophy and method that emphasizes the importance of selecting the right problem and the variables that most merit investigation. The highlights of the chapter are his taxonomy of scientific problems, the discussion of the scope of scientific principles, and the importance of having good enemies.

Nakayama, K. (1998). Vision fin-de-siècle - a reductionistic explanation of perception for the 21st century? In Hochberg, J., editor, *Perception and Cognition at Century's End: History, Philosophy, Theory*, pages 307–331. Academic Press.

von Békésy, G. (1960). *Experiments in Hearing*, chapter 1: The Problems of Auditory Research, pages 3–10. McGraw-Hill.

## 1.4 Supplemental Readings

Barlow, H. B. (1994). What is the computational goal of neocortex? In Koch, C. and Davis, J. L., editors, *Large-scale neuronal theories of the brain*, pages 1–22. MIT Press.

Churchland, P. S., Ramachandran, V. S., and Sejnowski, T. J. (1994). A critique of pure vision. In Koch, C. and Davis, J. L., editors, *Large-scale neuronal theories of the brain*, pages 23–60. MIT Press.

Marr, D. (1982). *Vision*, chapter 1: The Phiosophy and the Approach, pages 8–38. Freeman.

Richards, W. (1988). The approach. In Richards, W., editor, *Natural Computation*, pages 3–13. MIT Press.

# 2 Sound Localization

## 2.1 Topics

Lecture 1:

- human performance and sound localization psychophysics, minimum audible angle
- physics of sound propogation: decibel scale, reflection, interference, and acoustic shadowing
- simplifying the problem: lateralization
- ITDs, IIDs or ILDs, the duplex theory and its limitations
- computing ITD: convolution
- relating acoustic cues and available information to human performance

Lecture 2:

- phase ambiguity
- computation of ITDs in the auditory system
- monaural localization
- externalization
- the funciton of the pinnae, HRTFs
- auditory distance cues
- the role of head movements

Lecture 3:

- Discussion of readings
- the computational problem of sound localization
- types of computational models
- modeling laterization using linear systems

- inverting the model

### 2.2 Handouts

Chapter 3 of Yost gives an introduction to the physics of sound propagation. The first part of chapter 1 of Moore gives an introduction to sound and its frequency domain representation. Chapter 6 gives an introduction to sound localization and spatial hearing, covering many of the topics in the lectures. Chapters 9 and 10 of Castleman provide mathematical background on linear systems and the Fourier transform responsectively. These chapters will also serve as reference material for other topics later in the course.

Castleman, K. R. (1996). *Digital image processing*. Prentice Hall.
Chapter 9: Linear Systems Theory
Chapter 10: The Fourier Transform

Moore, B. C. J. (1997). *An Introduction to the Psychology of Hearing*. Academic Press, fourth edition.
Chapter 1: The Nature of Sound and the Structure and Function of the Auditory System
Chapter 6: Space Perception

Yost, W. A. (2000). *Fundamentals of Hearing*. Academic Press, fourth edition.
Chapter 3: Sound Transmission

Tutorial: Filtering, convolution, and the Fourier domain using Matlab.

### 2.3 Readings

King, A. J., Schnupp, J. W. H., and Doubell, T. P. (2001). The shape of ears to come: dynamic coding of auditory space. *Trends Cogn. Sci.*, 5:261–270.

Kulkarni, A. and Colburn, H. S. (1998). Role of spectral detail in sound-source localization. *Nature*, 396:747–749.

Semple, M. N. (1998). Auditory perception - sounds in a virtual world. *Nature*, 396:721–724.

## 2.4 Supplemental Readings

Hartmann, W. M. (1999). How we localize sound. *Physics Today*, pages 24–29.

# 3 Auditory Sensory Coding

## 3.1 Topics

Lecture 1:

- scientific method: experimental vs theoretical
- the computational problems of sensory coding
- functions of the outer and middle ear
- basilar membrane and traveling waves
- sound transduction
- coding of auditory nerve fibers: frequency tuning, phase locking, probabilistic firing
- modeling of the cochlear response
- coding large dynamic range
- response non-linearities

Lecture 2:

- the computational problem of auditory coding
- bottom-up approach: model cochlear processes
- cochleagrams
- top-down approach: derive codes from principles
- linear superposition and basis function representation
- information theoretic formulation of coding problem
- efficient coding of natural sounds
- time-frequency representations

## 3.2 Handouts

The second part of chapter 1 of Moore gives an introduction to how sound is represented in the auditory system (This chapter was handed out in an earlier lecture).

Moore, B. C. J. (1997). *An Introduction to the Psychology of Hearing*. Academic Press, fourth edition.
Chapter 1: The Nature of Sound and the Structure and Function of the Auditory System

### 3.3 Readings

Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nat. Neurosci.*, 5:356–363.

Olshausen, B. A. and O'Connor, K. N. (2002). A new window on sound. *Nature Neuroscience*, 5:292–294.

### 3.4 Supplemental Readings

Shamma, S. (2001). On the role of space and time in auditory processing. *Trends in Cognitive Sciences*, 5:340–348.

## 4 Visual Sensory Coding

### 4.1 Topics

Lecture 1:

- Optics: point spread function, focus and diffraction blur
- optical transfer function, 2D aliasing
- scotopic and photopic vision, dynamic range considerations
- photo transduction of color, chromatic aberration, modulation transfer function, cone spatial mosaic
- retinal motion

Lecture 2:

- The multitude of retinal output pathways
- information theory, entropy, capacity, redundancy
- visual coding of spatial structure, efficient coding theory
- basis function representation
- probabilistic models of visual structure, principal and independent component analysis (PCA and ICA), modeling non-Gaussian distributions

- efficient coding of natural images

Lecture 3:
- the organization and complexity of retinal circuity
- retinal ganglion cell receptive fields
- contrast sensitivity function, non-linear dependence on luminance
- information transfer over a noisy channel
- joint, conditional, and relative entropy, mutual information
- efficient coding by the fly contrast response function
- information theoretic models of non-linear contrast sensitivity
- natural image statistics, whitening

Lecture 4:
- adaptive, non-linear coding of intensity time series
- signal to noise ratio, coherence
- upper and lower bounds on channel capacity
- static non-linearities, adaptive gain control
- block coding of images, jpeg compression
- multiresolution representations and subband coding schemes
- time frequency space
- wavelets

## 4.2 Handouts

Wandell, B. A. (1995). *Foundations of vision*. Sinauer, Sunderland, MA.
Chapter 2: Image Formation
Chapter 3: The Photoreceptor Mosaic
Chapter 5: The Retinal Representation
Chapter 8: Multiresolution Image Representations

## 4.3 Readings

Atick, J. J. (1992). Could information-theory provide an ecological theory of sensory processing. *Netw. Comput. Neural Syst.*, 3:213–251.

Lewicki, M. S. and Olshausen, B. A. (1999). Probabilistic framework for the adaptation and comparison of image codes. *J. Opt. Soc. Am. A*, 16:1587–1601.

Olshausen, B. A. and Field, D. J. (2000). Vision and the coding of natural images. *American Scientist*, 88:238–245.

## 4.4 Supplemental Readings

Atick, J. J. and Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural Comput.*, 2:308–320.

Bialek, W. (2002). Thinking about the brain. In Flyvbjerg, H., Jülicher, F., Ormos, P., and David, F., editors, *Physics of Biomolecules and Cells*. Springer-Verlag, Berlin.
Section 4: Toward a general principle?

Borst, A. and Theunissen, F. E. (1999). Information theory and neural coding. *Nat. Neurosci.*, 2:947–957.

Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609.

Simoncelli, E. P. and Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24:1193–1216.

van Hateren, J. H. and Snippe, H. P. (2001). Information theoretical evaluation of parametric models of gain control in blowfly photoreceptor cells. *Vision Research*, 41:1851–1865.

# 5 Visual Motion

## 5.1 Topics

Lecture 1:

- the problem of what to compute, motion and optical flow fields, general vs specialized algorithms
- spatio-temporal contrast sensitivity, space-time separability, non-linear dependence on luminance
- representation of motion in x-y-t space, motion sampling and aliasing, spatio-temporal filtering
- local motion ambiguity and the motion "aperture" problem, plaids, integration of local motion cues, Weiss-Adelson gelatinous ellipses

- higher-order motion: self-motion, focus of expansion, biological motion, stereokinetic depth, second-order motion, implicit figure motion, shadow motion

Lecture 2:
- Differential techniques for estimating optical flow, motion gradient equation, motion constraint lines
- ill-posed problems, regularization and smoothing
- Lucas and Kanade, motion energy methods, multiscale methods
- limitations and difficulties
- noise considerations

## 5.2 Handouts

Horn, B. K. P. (1986). *Robot Vision*. MIT Press, Cambridge, MA.
Chapter 10: Motion Field and Optical Flow

Wandell, B. A. (1995). *Foundations of vision*. Sinauer, Sunderland, MA.
Chapter 10: Motion and Depth

## 5.3 Supplemental Readings

Poggio, T., Torre, V., and Koch, C. (1985). Computational vision and regularization theory. *Nature*, 317:314–319.

# 6 Perceptual Inference and Bayesian Modeling

## 6.1 Topics

Lecture 1:
- integrating local motion cues, plaid motion patterns
- motion models: intersection of contraints, vector averaging, object tracking, predictions of different models
- Bayesian inference using multiple information sources
- Weiss and Adelson: gelatenous ellipse illusions
- Bayesian estimation of velocity for single point, a generalized model for multiple points

- predictions of model and comparison to performance of human subjects

Lecture 2:

- higher-order ambiguity of edges, edges in real images
- directly observable image structure and intrinsic image structure
- vision as inverse graphics, perception as Bayesian inference
- probability as uncertainty, generative models
- Bayesian inference, likelihood, prior, and posterior distributions
- regularization as Bayesian inference

Lecture 3:

- Bayesian denoising and restoration
- using high-level knowledge to fill in missing information
- restoration of time-varying signals
- hierarchical Bayesian belief networks
- learning and inference in hierarchical statistical models
- posterior maximization and Gibbs sampling
- mathematical contributions of bottom-up and top-down information
- application to hierarchical learning and inference problems

## 6.2 Handouts

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman and Hall/CRC.
Chapter 1: Fundamentals of Bayesian inference
Chapter 2: Single-parameter models
Chapter 3: Introduction to multi-parameter models

## 6.3 Readings

Kersten, D. and Yuille, A. (2003). Bayesian models of object perception. *Current Opinion in Neurobiology*, 13:150–158.

Weiss, Y., Simoncelli, E. P., and Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nat. Neurosci.*, 5:598–604.

## 6.4 Supplemental Readings

Bialek, W. (2002). Thinking about the brain. In Flyvbjerg, H., Jülicher, F., Ormos, P., and David, F., editors, *Physics of Biomolecules and Cells*. Springer-Verlag, Berlin.
Section 3: Optimal performance at more complex tasks

Lewicki, M. S. and Sejnowski, T. J. (1997). Bayesian unsupervised learning of higher order structure. In *Advances in Neural Information Processing Systems*, volume 9, pages 529–535. MIT Press.

Poggio, T., Torre, V., and Koch, C. (1985). Computational vision and regularization theory. *Nature*, 317:314–319.

Simoncelli, E. P. (1999). Bayesian multi-scale differential optical flow. In Jähne, B., Haussecker, H., and Geissler, P., editors, *Handbook of Computer Vision and Applications*, pages 397–422. Academic Press.

Weiss, Y. and Adelson, E. H. (2000). Adventures with gelatinous ellipses - constraints on models of human motion analysis. *Perception*, 29:543–566.

# 7 Representation of Shape and Surfaces

## 7.1 Topics

Lecture 1:

- forms of intrinsic visual structure, complex nature of visual surface structure
- artificial vision problems, mid-level perception, choosing relevant structure
- approaches to computing shape, shape from shading
- luminance, irradiance, reflectance
- lightness equation, forshortening, BRDF (bidirectional reflectance distribution function)
- idealized surfaces, Lambertian surfaces, albedo
- surface normals, reflectance maps

Lecture 2:

- two-tone images, Mooney faces, top-down influences in shape perception
- bottom-up approaches to shape inference, volumetric primitives, line contours and junctions

- 3D structure ambiguities of gray scale and two-tone luminance contours
- two-tone image perception, effect of light source information

Lecture 3:

- Bayesian modeling of the generic viewpoint assumption (Freeman)
- probabilistic models of visual surfaces
- marginalization in probability theory
- Laplace's approximation
- surface roughness, reflectance, and specularity
- performance of shape from shading algorithms
- energy minimization and regularization approaches
- consistency constraints for shape from shading

## 7.2 Readings

Adelson, E. H. (2000). Lightness perception and lightness illusions. In Gazzaniga, M., editor, *The New Cognitive Neurosciences*, pages 339–351. MIT Press, Cambridge.

Freeman, W. T. (1994). The generic viewpoint assumption in a framework for visualperception. *Nature*, 368:542–545.

Mingolla, E. and Todd, J. T. (1986). Perception of solid shape from shading. *Biol. Cybern.*, 53:137–151.

Moore, C. and Cavanagh, P. (1998). Recovery of 3d volume from 2-tone images of novel objects. *Cognition*, 67:45–71.

## 7.3 Supplemental Readings

Freeman, W. T. (1996). Exploiting the generic viewpoint assumption. *Int. J. Comput. Vis.*, 20:243–261.

Forsyth, D. A. and Ponce, J. (2003). *Computer Vision: A Modern Approach*. Prentice Hall.
Chapter 4: Radiometry - Measuring Light
Chapter 5: Sources, Shadows, and Shading

Mamassian, P., Knill, D. C., and Kersten, D. (1998). The perception of cast shadows. *Trends in Cognitive Sciences*, 2:288–295.

Zhang, R., Tsai, P. S., Cryer, J. E., and Shah, M. (1999). Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:690–706.

# 8 Perceptual Constancy

## 8.1 Topics

Lecture 1:

- Higher-level dependency of perceptual constancies
- examples perceptual constancy, lightness constancy, brightness contrast effect
- object invariance to cast shadows
- size constancy, texture gradients
- shape and viewpoint constancy
- canonical reference frames and contex effects in face perception
- shape representation with templates, eigenfaces
- eigenface distortions and abstract shape representation

Lecture 2:

- limitations of template-based shape recognition
- Fourier spectra as invariant features
- linear combinations of views
- invariant feature maps by pooling
- Are enough features sufficient: Seemore (Mel, 1997)
- A computational model of attention and object-centered reference frames (Olshausen et al, 1993).

## 8.2 Readings

Olshausen, B. A., Anderson, C. H., and Van Essen, D. C. (1993). A neurobiological model of visual-attention and invariant pattern-recognition based on dynamic routing of information. *Journal of Neuroscience*, 13:4700–4719.

Purves, D., Lotto, R. B., and Nundy, S. (2002). Why we see what we do - a probabilistic strategy based on past experience explains the remarkable difference between what we see and physical reality. *Am. Scientist*, 90:236–243.

## 8.3 Supplemental Readings

Olshausen, B. A., Anderson, C. H., and Van Essen, D. C. (1995). A multiscale dynamic routing circuit for forming size-invariant and position-invariant object representations. *Journal of Computational Neuroscience*, 2:45–62.

Purves, D., Lotto, R. B., Williams, S. M., Nundy, S., and Yang, Z. Y. (2001). Why we see things the way we do: evidence for a wholly empirical strategy of vision. *Philos. Trans. R. Soc. Lond. B*, 356:285–297.

# 9 Auditory Structure

## 9.1 Topics

Lecture 1:

- Computational goals in audition
- review of sound representation
- higher-level auditory response properties
- perceptual constancy in audition
- reverberation
- the precedence effect
- dereverberation models
- the problems in auditory stream analysis
- auditory grouping cues

Lecture 2:

- auditory representation of pitch and timbre
- definitions and properties of pitch and timbre
- place theories of pitch perception
- periodicity and correlogram models of pitch perception

## 9.2 Readings

Shamma, S. (2001). On the role of space and time in auditory processing. *Trends in Cognitive Sciences*, 5:340–348.

Slaney, M. and Lyon, R. F. (1993). On the importance of time - a temporal representation of sound. In Cooke, M., Beet, S., and Crawford, M., editors, *Visual Representations of Speech Signals*, pages 95–116. John Wiley & Sons.

## 9.3 Supplemental Readings

Evans, E. F. (1992). Auditory processing of complex sounds - an overview. *Philosophical Transactions of the Royal Society of London B*, 336:295–306.

Shamma, S. (2003). Encoding sound timbre in the auditory system. *Iete Journal of Research*, 49:145–156.

Shamma, S. and Klein, D. (2000). The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. *Journal of the Acoustical Society of America*, 107:2631–2644.

# 10 Auditory Scene Analysis

## 10.1 Topics

Lecture 1:

- low- and high-level cues in auditory scene analysis
- cues for auditory grouping
- audio demonstrations of grouping cues
- modeling the cocktail party problem: blind source separation
- source amplitude distributions
- modeling non-Gaussian distributions
- connection to independent component analysis (ICA)

Lecture 2:

- computational auditory scene analysis
- limitations of linear blind source separation
- Cooke's model:
- modeling the auditory periphery

- extraction of dominant frequency
- grouping frequency channels
- modeling auditory scene exploration
- hierarchical view of auditory scene analysis
- duplex perception
- implementation of grouping principles

### 10.2 Readings

Cooke, M. and Ellis, D. P. W. (2001). The auditory organization of speech and other sources in listeners and computational models. *Speech Communication*, 35:141–177.

Mellinger, D. K. and Mont-Reynaud, B. M. (1996). Scene analysis. In Hawkins, H., McMullen, T., Popper, A., and Fay, R., editors, *Auditory Computation*, pages 271–331. Springer, New York.

### 10.3 Supplemental Readings

Hyvarinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. Wiley Interscience, New York.
Chapter 1: Introduction

## 11 Eye Movements and Visual Scene Analysis

### 11.1 Topics

Lecture 1:

- visual processing time
- types of eye movements
- nystagmus and visualizing retinal eye movements
- eye movements relative to foveal size
- eye movements during natural motion
- speed and accuracy of saccadic eye movements
- biological specialization in eye movements
- major functions in the control of eye movements

- visual jitter
- scene analysis and eye movements
- saccadic integration and suppression
- change blindness

Lecture 2:

- Modeling holistic properties of visual scenes
- visual memory and speed of visual processing experiments
- speed and flow of information in the visual system
- importance of scene context in object perception
- modeling the shape of the scene
- statistical regularities in natural scenes
- statistical characterization of scene categories

## 11.2 Readings

Oliva, A. and Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42:145–175.

## 11.3 Supplemental Readings

Land, M. F. and Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41:3559–3565.

Liversedge, S. P. and Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4:6–14.

Rensink, R. A. (2000). Seeing, sensing, and scrutinizing. *Vision Research*, 40:1469–1487.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, 53:245–277.

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in Cognitive Sciences*, 4:147–155.

Torralba, A. and Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, 14:391–412.

# 12 Perceptual Organization in Vision

## 12.1 Topics

Lecture 1:

- classic problems in visual perceptual organization
- perceptual organization in auditory scene analysis
- classic principles of grouping
- perceptual organization and visual cues
- attention and visual scene analysis
- Olshausen's dynamic routing model
- organizing information in a visual scene
- selective attention and saliency maps
- Itti and Koch's computational model
- Palmer and Rock's theory of perceptual organization
- Malik and Perona's model of texture segregation
- texture segmentation using local image models
- ICA mixture models and clustering

Lecture 2:

- grouping more abstract features
- shape constancy and grouping
- grouping by color, proximity, and common motion
- grouping and higher-level (top-down) knowledge
- grouping in natural images
- other grouping-like problems: cast shadows, transparency
- Camouflage
- How is top-down knowledge learned?
- Mark Brady's thesis on learning grouping: digital embryos

## 12.2 Readings

Itti, L. and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40:1489–1506.

## 12.3 Supplemental Readings

Guo, C. E., Zhu, S. C., and Wu, Y. N. (2003). Modeling visual patterns by integrating descriptive and generative methods. *International Journal of Computer Vision*, 53:5–29.

Itti, L. and Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2:194–203.

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 20:1254–1259.

Zhu, S. C. (2003). Statistical modeling and conceptualization of visual patterns. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 25:691–712.

# 13 Object Representation and Recognition

## 13.1 Topics

Lecture 1:

- computational problems in object recognition
- psychological theories of object recognition
- limitations of templates
- learning features for face recognition
- limitations of human-generated face features
- limitations of parts-based strategies
- object recognition and viewpoint invariance
- object recognition by invariant feature matching
- human performance on viewpoint invariance
- recognition of unseen objects
- object recognition by structural description
- classical approches to recognizing object shape
- structural description theories
- object recognition in young toddlers

## 13.2 Readings

Mel, B. W. (1997). Seemore: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Comput.*, 9:777–804.

Sinha, P. (2002). Recognizing complex patterns. *Nature Neuroscience*, 5:1093–1097.

## 13.3 Supplemental Readings

Edelman, S. (1997). Computational theories of object recognition. *Trends in Cognitive Sciences*, 1(8):296–303.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, pages 1150–1157.