Scene Analysis

- **Visual scene analysis:**
  - World: 3-D Objects arranged in 3-D space
  - Optical projection: Objects, background shmooshed together in 2-D array of pixels.
  - Visual perception: Recover the objects (which patches belong together), arrangements.
The Gestalt Laws of Grouping

The Gestalt Law of “Common Fate”
Figure-Ground Segregation

Illusory Contours
9.00 Introduction to Psychology
Professor S. Pinker
Week 5, Lecture 2: Visual and Auditory Scene Analysis
Top-Down Influences on Scene Analysis

The Problem of Shape Recognition
Three Better Theories of Shape Recognition

1. Multiple templates
2. Mental transformations (imagery)
3. Object-centered reference frames (geons)

Visual Frames of Reference (Coordinate Systems)

- Viewer-centered frame: X aligned with retinal left-right axis; Y aligned with retinal up-down axis; Z aligned with line of sight
- Object-centered frame: X aligned with an axis of the object: elongation, symmetry, intrinsic front-back. Other axes at right angles to x.
The problem with using a viewer-centered reference frame:

The advantage of using an object-centered reference frame:

The Effects of Reference Frames on Shape Perception:

Effects of the Up-Down Axis:

Sometimes the Brain aligns Reference Frames with Objects, not just the vertical axis:
Testing the Three Theories  
(Tarr & Pinker, 1989, 1990)

- If people use object-centered representations, they should recognize an object equally well at all orientations
- If people use mental rotation, they should take longer for objects that are more tilted
- If people use multiple templates, they should take longer for orientations they haven’t seen before
Asymmetrical objects are mentally rotated:

Symmetrical objects are matched in an object-centered reference frame:

Train at some orientations, test at others:

People are faster with orientations they have been trained on:
Visual and Auditory Scene Analysis

- **Visual scene analysis:**
  - World: 3-D Objects arranged in 3-D space
  - Optical projection: Objects, background shmooshed together in 2-D array of pixels.
  - Visual perception: Recover the objects (which patches belong together), arrangements.

Auditory Scene Analysis

- World: Things making noise (sound sources) -- each produces a set of frequencies.

  - Wave superposition: all frequencies shmooshed together in complex waveform.
Auditory Scene Analysis:

- World: Things making noise (sound sources) -- each produces a set of frequencies.
- Wave superposition: all frequencies shmooshed together in complex waveform.
- Auditory perception: Recover the soundmakers (which frequencies belong together), arrangement.

Scene Analysis, continued

- Frequency analysis by the ear and brain.
How Auditory Scene analysis is similar to Visual Scene Analysis:

- Space in vision = frequency X time in audition.
- Object in vision = sound source (noisemaker) in audition.
- Gestalt laws of auditory organization: proximity, continuation, common fate (motion).

Psychophysics of Sound: Frequency is perceived as pitch

frequency

time

Auditory Streaming

<table>
<thead>
<tr>
<th>B</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>
Gestalt laws reflect physical constraints:
- In vision, matter is cohesive and bounded: parts of an object stick and move together
- In audition, soundmakers have resonances and nonzero mass: they can’t change pitch instantaneously

Simultaneous & Sequential Grouping

Gestalt Continuity in Audition

Auditory Scene Analysis & Music

Segregation of high notes from low ones in a sonata by Telemann.
Streaming in African xylophone music.