Visual Object Recognition

Neurobiology 230 – Harvard / GSAS 78454

Gabriel Kreiman

Email: gabriel.kreiman@tch.harvard.edu
Phone: 617-919-2530
Web site: http://tinyurl.com/vision-class
Dates: Mondays
Time: 3:30 – 5:30 PM
Location: Biolabs 1075
Psychophysics: The study of the dependencies of psychological experiences upon the physical stimuli that generate them

Basic measures:

• Reaction time — The time taken by subjects to perform a task or make a judgment can give an indication (or at least an upper bound) of how long the necessary psychological (and hence neural) processing takes.

• Performance — Often inversely related to reaction time. There are techniques for mitigating response biases.

• Threshold — Stimuli can be varied to determine the threshold for detection, discrimination, or some more complex psychological phenomenon.
Gestalt laws of grouping
Basic phenomenological constraints

- **Law of Closure** — The mind may experience elements it does not perceive through sensation, in order to complete a regular figure (that is, to increase regularity).
- **Law of Similarity** — The mind groups similar elements into collective entities or totalities. This similarity might depend on relationships of form, color, size, or brightness.
- **Law of Proximity** — Spatial or temporal proximity of elements may induce the mind to perceive a collective or totality.
- **Law of Symmetry (Figure ground relationships)** — Symmetrical images are perceived collectively, even in spite of distance.
- **Law of Continuity** — The mind continues visual, auditory, and kinetic patterns.
- **Law of Common Fate** — Elements with the same moving direction are perceived as a collective or unit.
Law of closure: perceiving objects as whole even if they are not complete.

The mind may experience elements it does not perceive through sensation, in order to complete a regular figure (that is, to increase regularity).
Law of similarity

The mind groups similar elements into collective entities or totalities. This similarity might depend on relationships of form, color, size, or brightness.
Law of proximity

• Spatial or temporal proximity of elements may induce the mind to perceive a collective or totality.
Law of continuity

The mind continues visual, auditory, and kinetic patterns
Law of common fate
Holistic representation of faces

A Thatcher illusion

McKone et al, Frontiers in Psychology, 2013
Holistic representation of faces

Composite illusion

McKone et al, Frontiers in Psychology, 2013
Holistic representation of faces

McKone et al, Frontiers in Psychology, 2013
Tolerance to image transformations

- Scale
- Position
- Rotation (2D)
- Rotation (3D) – viewpoint
- Color
- Illumination
- Cues
- Clutter
- Occlusion
- Other non-rigid transformations (aging, expressions, etc)
Scale tolerance
One-shot learning for scale tolerance

Which one is it?
Position tolerance
Other transformations require example-based training
Tolerance to viewpoint and illumination changes
Recognition from minimal features
Recognition of caricatures

Images: Hanoch Piven
Beyond pixels – Context matters
Visual recognition depends on experience
Recognition of images flashed for ~100 ms (demo)
Visual recognition can be extremely fast


Table 1
Summary of behavioural results. Participant numbers correspond to those in Fig. 4.

<table>
<thead>
<tr>
<th>Subject</th>
<th>N</th>
<th>Accuracy (%)</th>
<th>Median RT (ms)</th>
<th>Min RT (ms)</th>
<th>Mean start (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>682</td>
<td>96.3</td>
<td>227</td>
<td>130</td>
<td>143</td>
</tr>
<tr>
<td>2</td>
<td>774</td>
<td>93.3</td>
<td>200</td>
<td>130</td>
<td>136</td>
</tr>
<tr>
<td>3</td>
<td>726</td>
<td>81.8</td>
<td>201</td>
<td>130</td>
<td>129</td>
</tr>
<tr>
<td>4</td>
<td>563</td>
<td>80.1</td>
<td>191</td>
<td>120</td>
<td>126</td>
</tr>
<tr>
<td>5</td>
<td>672</td>
<td>86.6</td>
<td>159</td>
<td>130</td>
<td>133</td>
</tr>
<tr>
<td>6</td>
<td>675</td>
<td>86.1</td>
<td>224</td>
<td>150</td>
<td>143</td>
</tr>
<tr>
<td>7</td>
<td>574</td>
<td>90.2</td>
<td>204</td>
<td>140</td>
<td>129</td>
</tr>
<tr>
<td>8</td>
<td>653</td>
<td>94.0</td>
<td>213</td>
<td>150</td>
<td>147</td>
</tr>
<tr>
<td>9</td>
<td>694</td>
<td>96.7</td>
<td>251</td>
<td>180</td>
<td>200</td>
</tr>
<tr>
<td>10</td>
<td>534</td>
<td>89.7</td>
<td>236</td>
<td>180</td>
<td>124</td>
</tr>
<tr>
<td>11</td>
<td>739</td>
<td>90.0</td>
<td>253</td>
<td>190</td>
<td>205</td>
</tr>
<tr>
<td>12</td>
<td>652</td>
<td>96.6</td>
<td>276</td>
<td>200</td>
<td>235</td>
</tr>
<tr>
<td>13</td>
<td>703</td>
<td>95.0</td>
<td>238</td>
<td>160</td>
<td>173</td>
</tr>
<tr>
<td>14</td>
<td>769</td>
<td>98.7</td>
<td>301</td>
<td>230</td>
<td>251</td>
</tr>
<tr>
<td>15</td>
<td>529</td>
<td>77.1</td>
<td>233</td>
<td>160</td>
<td>235</td>
</tr>
<tr>
<td>All</td>
<td>8998</td>
<td>90.1</td>
<td>228</td>
<td>120</td>
<td>140</td>
</tr>
</tbody>
</table>

The second column of this table indicates the total number of trials per participant (see Section 2 for details). Columns 3–5 give the mean accuracy, median and minimum reaction time values for each participant shown in Figs. 3B and C. The last column indicates the onset latency of the mean eye trace for each participant (see Fig. 5).
Backward masking

10 ms  20 ms  30 ms  40 ms  50 ms  100 ms  200 ms
Doubles?

http://www.francoisbrunelle.com/

Francois Brunelle
Is information integrated over time?

Original image
Atom 1
Atom 2
Atom 3

Which category?
Animal
Person
Plant
Vehicle

Time
SOA
SOA

500 ms fixation with eye tracking
500 ms of 170 Hz noise
500 ms noise, then response screen

Singer and Kreiman, 2014
Rapid decay in recognition of asynchronously presented object parts

Brief asynchronies disrupt object recognition

Singer and Kreiman, 2014
The visual system has a very large capacity
A massive recollection capacity

Pattern completion: Objects can be recognized from partial information.
Amodal completion
Object recognition from partial information
Object completion task

Unmasked

Whole

Animal
Chair
Vehicle
Fruit
Face

response

500 ms

x

33-150 ms

500 ms

Partial

Animal
Chair
Vehicle
Fruit
Face

response

500 ms

x

33-150 ms

500 ms
Object completion (unmasked condition)

Whole

![Graph showing performance vs SOA for whole images.]

Partial

![Graph showing performance vs SOA for partial images.]

Unmasked

![Graphs showing response times for whole and partial images with unmasked conditions.]

Partial

![Graphs showing response times for whole and partial images with partial conditions.]

Legend:
- Animal
- Chair
- Vehicle
- Fruit
- Face

Response times:
- 33-150 ms
- 500 ms
Object completion task (masking)
Object completion (unmasked condition)

Unmasked

Masked
Further reading


Original articles cited in class (see lecture notes for complete list)

- McKone et al, Frontiers in Psychology, 2013