Perceptual Processes I: Visual and Auditory Recognition

Chapter 2

Perception: An Introduction

Perception is a mental (cognitive) process that involves using previous knowledge to gather and interpret the stimuli gathered by the senses. Thus perception involves input and recognition of sensory experience.

Sensation refers to activation of the receptors in the sense organs by stimulation. Salt can activate a taste bud. This input of the stimulus is sensory experience.

Background on Visual Object Recognition
Object Recognition

During object recognition you identify arrangement of sensory stimuli and perceive a pattern separate from background.

Perceptual Stimuli

1. Psychologists talk about two kinds of perceptual stimuli. Distal stimulus (the object) and the proximal stimulus (the retinal image).

2. When an object is recognized, the identity of the distal stimulus is determined, even though the information in the proximal stimulus is not perfect (Kersten et al., 2004; Pasternak et al., 2003).

Visual System

3. Retinal image (proximal stimulus) then moves thorough the visual pathway and reaches the primary visual cortex (and other cortical areas) to be processed for perception.
Perceptual Framework

<table>
<thead>
<tr>
<th>Sense</th>
<th>Distal Object</th>
<th>Medium</th>
<th>Proximal Stimulation</th>
<th>Perceptual Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vision</td>
<td>Pencil</td>
<td>Light</td>
<td>Photons hitting receptors</td>
<td>Pencil</td>
</tr>
<tr>
<td>Audition</td>
<td>Thunder</td>
<td>Sound</td>
<td>Waves hitting hair cells</td>
<td>Thunder</td>
</tr>
<tr>
<td>Olfaction</td>
<td>Musk</td>
<td>Molecules of musk in air</td>
<td>Molecules absorbed by hair cells</td>
<td>Musk smell</td>
</tr>
<tr>
<td>Gustation</td>
<td>Strawberry</td>
<td>Molecules in air/water</td>
<td>Molecules absorbed by taste buds</td>
<td>Strawberry taste</td>
</tr>
<tr>
<td>Touch</td>
<td>Sandpaper</td>
<td>Silica granules on paper</td>
<td>Deformation of skin receptors</td>
<td>Surface Roughness</td>
</tr>
</tbody>
</table>

Adapted from Strenberg (2003)

Organization in Visual Perception

1. It takes less than 100ms (1/10 sec) to recognize an object. Iconic memory helps. But how do we recognize objects amidst a background?
2. Gestalt psychologists suggested that human beings have an innate tendency (law of Prägnanz) to organize visual images into patterns (figure), and process the rest as background.

Organization in Visual Perception

3. Figure (object) has definite shape (but see illusory contours below), seems closer and more dominant than the background. See reversible figure.
4. Explanation of reversible has two components. A. Neurons in the visual system become adapted to say the faces and then switch to the vase, and vice versa. B. We solve visual paradox problems by alternating between two reasonable solutions.
Organization in Visual Perception

6. We do perceive figure-ground relationships even when the scene has no clear cut boundary. This is observed in illusory contours.

7. The inverted triangle is perceived in front of the upright triangle and circles. The color of the triangle seems brighter than the background.

Illusory Contours

Organization in Visual Perception

8. We believe in our ability to perceive objects and deem them stable. Figures or objects remain perceptually constant even though our proximal stimulation may change.

9. Perceptual mechanisms constantly readjust proximal stimulation, to maintain perceptual constancy. So memory plays an important role compensating changes in confusing stimuli.

The door remain rectangular in our perception even when its views change.

Visual Object Recognition: Theories

Researchers have proposed a variety of theories to explain object recognition. Some of these theories are old and simple others new and complicated.

1. Template-matching theory suggests that our perceptual systems use templates in our memory for recognizing patterns and objects.
Template Matching Theory

2. Some machine recognition is based on template matching recognition scheme.

3. However, template matching theory breaks down when letters or objects, to be recognized, do not match the template.

Template Matching Theory

<table>
<thead>
<tr>
<th>Advantage</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>Rigid. Inflexible.</td>
</tr>
<tr>
<td></td>
<td>Unwieldy (many templates required).</td>
</tr>
<tr>
<td></td>
<td>Inaccurate (recognition based on partial templates).</td>
</tr>
<tr>
<td></td>
<td>Inefficient (Would be able to recognize fewer patterns or objects than there are in the environment).</td>
</tr>
</tbody>
</table>

Feature-Analysis Theory

1. Visual stimulus is composed of many features with distinctive forms.
2. Features of objects and patterns are extracted and matched to features in memory for recognition.
3. We can decompose the letter “A” into at least three features (/ - \) or line segments.
Feature-Analysis Theory

1. Arguments for feature-analysis theory come from reaction time studies.
2. There are more feature confusions for letter “Z” in list 1 than list 2, therefore takes longer subjects to find it, and they make more errors (Neisser, 1964).

Feature-Analysis Theory

Pandemonium Model

Selfridge (1959) model explains how features of a letter processes through stages to be perceived.
Neurobiological Evidence

Hubel & Wiesel (1960) provided neurobiological evidence of feature detectors (neurons) in the primary visual cortex.

Recognition-by-Components Model

1. Biederman (1987) suggested that we perceive objects by using what he refers to as "Geons" (short for geometric icons).
2. Geons are like 3D features.
3. He lists 36 Geons for 3-D recognition.

Geons and Objects

Geons are elements that form 3D objects.
Where do Geons come from?

How does our perceptual system take line segments and forms volumetric primitives (Geons)?

David Marr (1982) suggested a number of computational steps take place in the brain, before the object becomes a 3D representation.

Top-down Processing and Visual Object Recognition

Object Recognition

Top-down Theories

1. Top-down theories emphasize that recognition of objects is based on concepts, expectations, and memories. Also called conceptually driven processing or constructive perception.

2. The following phrase is read as “the cat” when in fact letter “H” and “A” are the same.

TAE CAT
Top-down Theories

3. We use context based memory to say there is coffee in the cup when posed with the question.

Bottom-up Theories

1. Bottom-up or data driven processing emphasize the importance of stimulus in object recognition (cf. J. J. Gibson, 1966). Gibson says object recognition is based on invariant properties of the object. Gordon (2004) disagrees and suggests that higher and more sophisticated levels in the perceptual systems are engaged for object perception.

2. In object recognition, the first part may be bottom-up processing but instantaneously top-down processing joins in (Palmer, 2002).

Features Processed?

If object recognition was only based on bottom-up processing, say in reading a text, we would be processing about 5000 features (of letters) per minute to read, which is staggeringly high. Therefore, a top-down processing (context driven) is inevitable, and processes information faster.
Read Aloud

According to a research at Cambridge University, it doesn't matter in what order the letters in a word are, the only important thing is that the first and the last letter be at the right place. The rest can be a total mess and you can still read it without problem. This is because the human mind does not read every letter by itself, but the word as a whole.

Rayner and colleagues (2006) found that the speed of reading text, like above, drops down slightly.

Word Superiority Effect

1. We can recognize a single letter more accurately and rapidly when it appears in a meaningful word than when it appears in a meaningless string of letters or if it appears alone (Palmer, 2002).
2. This demonstrates facilitation by top-down processing.

<table>
<thead>
<tr>
<th>Word</th>
<th>String</th>
<th>Letter</th>
</tr>
</thead>
<tbody>
<tr>
<td>WORK</td>
<td>ORWK</td>
<td>&quot;R&quot;</td>
</tr>
<tr>
<td>ISLAND</td>
<td>LASNDI</td>
<td>&quot;S&quot;</td>
</tr>
</tbody>
</table>

Reicher (1969)

Word-in-a-Sentence Effect

1. A similar effect is produced when we recognize a word in a sentence. Rueckl and Oden (1986) demonstrated that features of a stimulus and the nature of the context influences word recognition.
2. They manipulated the shape (features) of letters "n" and "r" and three intermediaries, and the words "beans" and "bears" in sentences.

beans bears
Word-in-a-Sentence Effect

3. And then constructed four sentences like following to manipulate the context.

1-2. The lion tamer (or zookeeper) raised beans to supplement his income. The likely response was bear.

3-4. The botinist (or dairy farmer) raised beans to supplement his income. The likely response was beans.

Results

Results suggest, when the letter was more like “r” (or the context was lion tamer and zookeeper) recalled word was bear, and vice versa.

Change Blindness

Change blindness refers to the inability to detect changes in an object or a scene (Simon & Levin, 1997) and is due to overactive top-down processing.
Inattentional Blindness

Inattentional blindness refers to the inability to pay attention to an object or individual in a scene. About half of the participants were blind to the man with the umbrella or in a gorilla suit.

http://www.youtube.com/watch?v=0AwwlJtnwA8
Face Recognition in Infancy

Babies nine minutes old spend more time looking at pictures that resemble human faces than jumbled faces, lines or other objects. This suggests perhaps we have innate ability to recognize faces.

Face vs. Object Recognition

1. Studies have suggested that face recognition may be different from how we recognize other objects or patterns.
2. We tend to recognize faces on holistic basis compared to other perceptual objects (Farah, 1996).

Learn

(Farah, 1996)
1. Much of the neuroscientific understanding about face recognition comes from lesion studies.
2. In a peculiar case an individual lost his ability to recognize human faces (prosopagnosia) and yet was able to recognize his sheep (McNeil and Warrington, 1993).
3. Prosopagnosia is a medical condition in which the individual is unable to recognize faces.
4. McNeil and Warrington (1993) report that prosopagnosia is due to damage in the fusiform gyrus of the brain. However, face recognition in the brain is likely to be represented in the inferotemporal cortex (Vecera, 1998).

Neuroscience & Face Recognition

5. Other agnosias (Agnosia [a “lack”, gnosis “knowledge”]) include being unable to recognize objects (Luria, 1973), or Simultagnosia: Inability to recognize multiple elements in a visual presentation except one. Tree but not forest.

Biology of Face Recognition

Human brain responds faster and more accurately to pictures of upright faces compared to inverted faces (fMRI, D’Esposito et al., 1999; McKone, 2004).
Applied Research

1. Applied research with face recognition has been carried out in a number of governmental and private organizations.
2. These have included credit cards, ID cards, passports other security identifications.

Recognize your Professors

1. Burton et al., (1999) tested three groups (familiar students, unfamiliar students, and police officers) in recognizing professors.
2. Familiar students had already known their professors, but the unfamiliar and police officers had seen professors only in videos.
3. Confidence ratings of familiar students was highest compared to unfamiliar students and police officers.
Face Recognition: Schizophrenia

1. Patients with schizophrenia have difficulty recognizing faces.
2. Martin et al., (2005) gave schizophrenics and controls to look at two photographs one after another and a) judge them as same or different, b) judge whether the faces depicted same or different emotions.
3. Schizophrenics took longer time to judge photos for facial content and made many errors. The same was true when they judged emotions.

Speech Perception
Speech Perception

1. During speech perception auditory system must record the sound vibrations by the talker and translate that to speech.
2. We process about 900 sounds (phonemes) per minute. Quite a staggering amount.
3. Separate each word from other words in memory.
4. Separate speech from non-speech sounds (noise).

Speech Perception: Characteristics

1. English language contains 44 phonemes, basic unit of spoken language such as a, k, and th etc. Phonemes consist of consonants and vowels.
2. Listeners can impose word boundaries in speech when words are not separated by silences.
3. Phoneme pronunciation varies tremendously.
4. Context allows listeners to fill in the missing sounds.
5. Visual cues from the speaker’s mouth help us interpret ambiguous sounds.

Word Boundaries

1. It seems we are unable to recognize word boundaries in an unfamiliar language. However we can define such boundaries in English quite easily.
2. Even pauses are less than 40% silent.
3. Research has shown that there is an inherent system that effortlessly divides the words into boundaries.
Phoneme Pronunciation

1. Phoneme perception is difficult. For example speakers differ in pronouncing phonemes by affecting their pitch, tone and rate.
2. A second factor that affects phonemes recognition is that at times they are omitted in (words) speech.
3. Phoneme articulation is affected by flaking phonemes. Saying t in “water” (t flanked by a and e) would be different from saying t in “function” (t flanked by c and i).

Context and Speech Perception

1. Context helps reconstruct phonetic sounds (phonetic restoration) and words. So top-down processing helps in reconstruction of speech.
2. Waren and Waren (1970) showed the effect of context on restoring a partial word.
   a. It was found that the *eel was on the axle.
   b. It was found that the *eel was on the shoe.
   c. It was found that the *eel was on the orange.
3. Appropriately words wheel, heel and peel were restored.

Visual Cues and Speech Perception

1. When speech is presented with white noise and we listen to it with our eyes closed... it is difficult to perceive speech. However if we open our eyes and look at speaker’s lips, speech perception is facilitated (Smyth, 1987).
2. People with normal hearing do not take full advantage of visual cues, except certain circumstances (Badly dubbed French movie).
3. McGurk effect. Lips say “gag” and the ears hear “bab”, speech perception ends up as a compromise “dad”.

---

55

56

57
Theories of Speech Perception

Some theorists believe that we have special brain mechanisms that explain our impressive skills. Others believe that speech skills are impressive but are no different from other cognitive functions.

1. **Speech-is-special approach**, suggests that humans are born with a special device that allows them to decode speech than non-speech (instrumental music).
2. These researchers talk about a phonetic module that handles speech perception but not other forms of auditory perception.

Speech-is-Special Approach

3. Speech-is-special approach, suggests that speech perception is organized in the brain in a special way. So speech is handled differently than other cognitive functions like memory, solving problems etc (Trout, 2001).
4. One argument that favors this approach emphasizes categorical perception. When presented with a phonetic sound between a b and p. People either perceive a “b” or “p” than the ambiguous sound.

General Mechanism Approaches

1. Most theorists now believe in a general mechanism approach which suggests that we perceive speech sounds similar to processes used in non-speech sounds.
2. Humans exhibit categorical perception for non-speech sounds also (Pastore et al., 1990).
3. Since visual cues facilitate speech perception argument for a phonetic module is weakened.