9.65 February 9, 2004 Object recognition HANDOUT

I. Why is object recognition difficult?
II. Basics of visual perception
    A. Gestalt principles, constancy
    B. Computational approaches: e.g., perception of
        surfaces
III. Higher-level vision: Object recognition
IV. Word perception

Today: Visual perception, leading to higher-level vision: object recognition, word perception.

I. Why is object recognition difficult?

For example, what's this?

This example illustrates the ambiguity of perception, and its interpretive nature.We combine information from our senses with our knowledge of what's likely, especially what's likely in context.

The early visual system extracts information that is likely to be correlated with actual objects in the environment--under these generally true assumptions about objects:

-continuous boundaries

-surfaces that change texture and color and brightness relatively smoothly, if the surface is continuous, but change sharply when the surface is discontinuous.

-move as a unit

-rigid (or they deform in a regular way).

Why simple solutions don't work in object recognition

SEPARATING OBJECTS IN A SCENE: difficult

TEMPLATE-MATCHING (once a single object has been segregated)--why this often won't work

-need to recognize in [almost] any position and rotational orientation

-any size on the retina

-any brightness

-any illuminant color

　　　-with various distortions or variations

Learn by experience with each specific object? To some extent, but we generalize well to shifts of size, orientation.

So: we must have some way of abstracting certain "invariants" from an object.

Biederman's RBC theory, in your text (Chap. 3).

II. Basics of visual perception:

A. Gestalt principles, perceptual constancies, the Weber-Fechner Law :

Gestalt psychologists: Wertheimer, Koffka, Ko"hler, and others.

The first principle: **figure-ground organization**:

figure "captures" the common contour, figure has more definite localization, more solid (less filmy) color

Other principles:

　proximity

　similarity

　good continuation

　closure: dotted lines joined

　common fate (moving together)

All represent "good form" [Law of Pra"gnanz]: simplicity, inclusiveness, symmetry,

regularity.

The perceptual constancies:

The Gestalt psychologists and their followers also emphasized the RELATIONAL character of perception: we see relative (not absolute) brightness, so that a white piece of paper still seems white in a dim room, when the light intensity reflected from it may be lower than that of a black piece of paper in a bright light.

This is perceptual CONSTANCY--perception "correcting" for the brightness or color of illumination, correcting for the angle at which we view something, the object's size on the retina as a function of distance, etc.

Perceptual constancy is not perfect, however: we do notice changes. Just how much change does there have to be, before we notice it?

The Weber-Fechner Law: Fechner's Law: $S = k \log I$, where S is the psychological sensation, I is the physical measure, e.g., intensity of light, and k is a constant that differs for different physical dimensions. This law is a good descriptive generalization, but it tends not to hold at the extremes of any dimension.

The perceived "strength" of a sensation such as the brightness of a light grows as a RATIO of the light's physical intensity to that of the light you are comparing it with.

Substituting a 100 watt bulb for a 50 watt bulb makes a very noticable difference, whereas going from 200 to 250 watts is much less noticeable.

This ratio law reflects a compromise between perfect constancy (which would leave us unable to detect a shadow, for example) and a perfect correlation between light energy and perceived brightness (which would make the same object look too extremely different as lighting changed).

To sum up, the Gestalt psychologists focused on the abstract, inferential character of perception: we don't "see" what our retina sees, but instead perception is the consequence of a complex set of processes.

B. Computational approaches to object perception:

A more unified, computational account of the many phenomena described by Gestaltists and other visual scientists.

David Marr

An important goal is to discover CONSTRAINTS or assumptions made by the visual system: to resolve AMBIGUITY.

One example: **perception of surfaces**.

Recommended: Nakayama, He, & Shimojo (1995), chapter in Kosslyn & Osherson (Eds.), **Visual cognition: Vol. 2, An invitation to cognitive science** 2nd Edition.

One illustration of this approach: the perception of figure and ground.

Principle One: only one surface can "own" a boundary between it and another surface.

Principle Two: we are built to see probable events, not unlikely ones.

Generic (typical) view

Accidental (coincidental) view.

Hard-wired and MODULAR?

III. Higher-level vision:

Role of TOP-DOWN CONTEXT

Demo:

IV. Word perception:

a. Words in isolation:

The Wheeler-Reicher WORD SUPERIORITY EFFECT (WSE) (see text, pp. 68-69):   It's easier to detect a letter in the context of a word than in isolation.

Experiment:

```
                          K
      _ _ _ K -->  # # # # --> _ _ _ _
                                    D
```

(which did you see: K or D?)

OR:

```
                              K
        W O R K --> # # # # --> _ _ _ _
                              D
```

-word shape doesn't matter: WORK, work, even WoRk

-the effect requires a very clear glimpse, followed by patterned masking--not dim or indistinct, followed by a blank field

-you still get some benefit from pronounceable "pseudowords": REET, MAVE as compared with random, nonpronounceable strings: RTEE, AEVM.

Explanations of these effects require that, at the brief moment of perception, possible words CONSTRAIN what letters are seen: tentative letter assignments constrain each other.

```
     C        C
   T   E    T   E     THE vs. CAT:
   T        T         H OR A? (Fig. 3.6 in text)
```

In Chapter 3 several examples of simple network models are described to give you an idea about how the visual system might place its bets effectively, even when reading rapidly or in imperfect visual conditions. The McCLELLAND-RUMELHART Interactive Activation Model (an influential early model) is also described.

b. Word perception in context

Normally we don't read words in isolation, but in sentence or paragraph CONTEXT.

PROOF-READING

Context experiment: Tulving, Mandler, & Baumal (1964) gave participants 0, 2, 4, or 8 words of sentence context (9 was full sentence),and then a very brief word for 0 to 140 ms. Ascending method: first saw word at 0, then increasing durations until recognized.

  E.g., An apple/from the/ tree hit/ my bare/ HEAD.

Results:

Viewers required a longer exposure duration to recognize final word, the shorter the preceding context.

Problems:

Subjects had multiple chances to guess the word, as the duration increased --unlike normal reading. And they could think about the context as long as they wished, before guessing the word.

OTHER EXPERIMENTS on context: Immediate measures:

Potter, Moryadas, Abrams, & Noel, 1993, used RSVP presentation of sentences, so that subjects didn't have time to puzzle out what each word was, in relation to the context. They showed that the context was still able to influence recognition of a critical word that might or might not belong in the context: e.g., horse/house/honse:

a. They looked at the horse from their car. *Neutral: horse or house usually read correctly*

b. The boy rode the house around the pasture. *Biased against word: house often misread as horse.*

c. The lawn in front of the honse was overgrown. *Nonword honse was often misread as house, rarely as horse.*

A McClelland-Rumelhart type of model could readily explain these results, but only if expanded to include high-level knowledge as a source of top-down input once several alternatives are proposed by visual analysis.

CONCLUSION ABOUT CONTEXT AND WORD PERCEPTION:

Initial access on basis of perception only? Context selects or confirms.