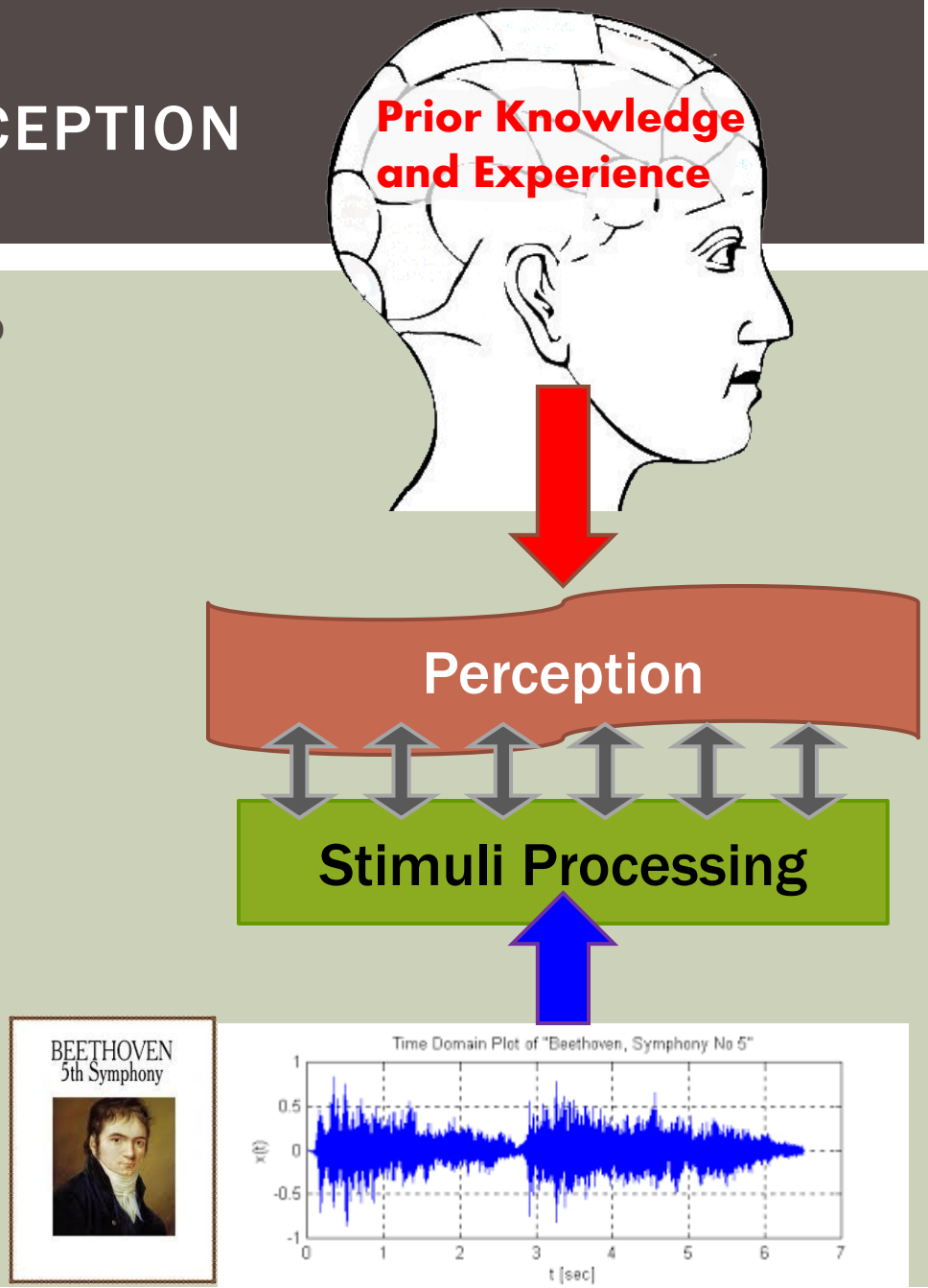
Perception

Most slides in this presentation were developed by Rowan professor Mark Hale. Professor Hale is a Cognitive Psychologist and Human Factors Specialist and very gratefully contributed his material to this course in HCI

Perception

- Perception uses previous knowledge to gather and interpret the stimuli registered by the senses.

- Notice that perception combines aspects of both the outside world (the stimuli) and your own inner world (your previous knowledge).



Some content courtesy of Professor Mark Hale

Human Perception is Amazing

- Our perceptual abilities are amazing!
  Yet, we generally take them for granted. Consider:

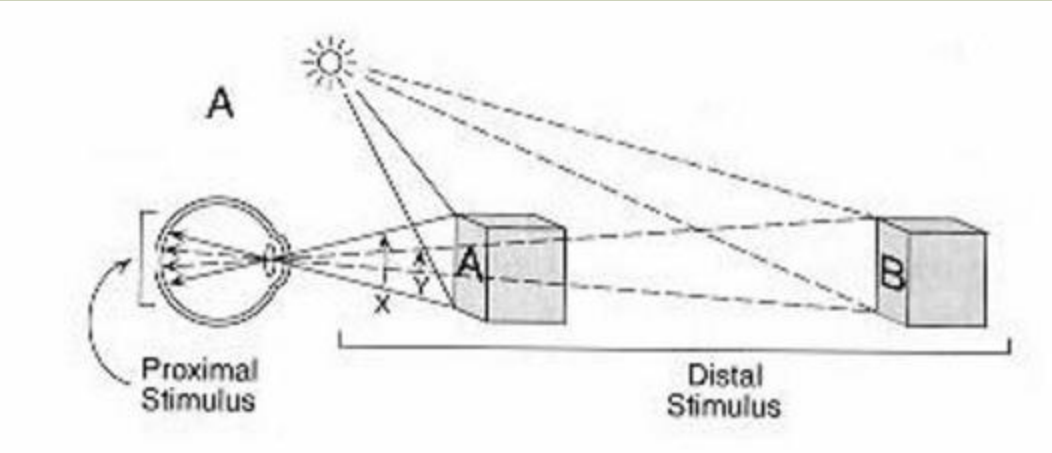
  - **Vision**: We can see a single candle flame from 30 miles on a clear night
  - **Hearing**: We can hear the tick of a watch from 20 feet in absolute silence
  - **Smell**: We can smell one drop of perfume in a 6-room apartment
  - **Touch**: We can feel the wing of a bee on our cheek, when dropped from 1 cm
  - **Taste**: We can taste one teaspoon of sugar in 2 gallons of water

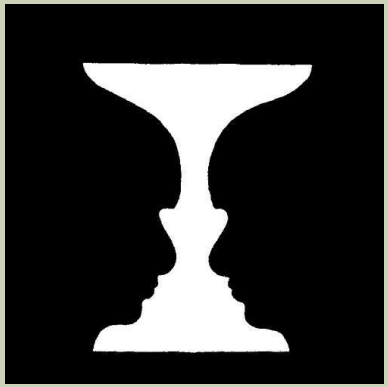
  These are all under ideal conditions where noise is minimal.

  - Glancing at this slide you can interpret hundreds of small squiggles into meaningful and coherent thoughts with relative ease.
  - Similarly we recognize spoken words with ease.
  - We can recognize faces instantly of people we haven't seen in a very long time.

Visual Object Recognition

- There are two types of perceptual stimuli.
  - Distal Stimulus – Actual object "out there"
  - Proximal Stimulus – The information registered on your sensory receptors.





Figure/Ground

- During object/pattern recognition, you identify a complex arrangement of sensory stimuli, and you perceive that this pattern is separate from its background (Figure/Ground).
- You then transform and organize the raw information provided by your sensory receptors and compare it to other information in memory storage.

Content courtesy of Professor Mark Hale

Visual Object Recognition

- When we recognize an object, we manage to figure out the identity of the distal stimulus, even when the information available in the proximal stimulus is far from perfect.

- Why might the proximal stimulus be a less than perfect representation of the distal stimulus?

- There are lots of reasons!
  - Things appear to be differently shaped depending upon viewing perspective – constancies (shape, size, volume)
  - The proximal stimulus is 2D while the distal is 3D.
  - Not only is the information far from perfect, our sensory gathering mechanisms (organs) are far from perfect.

Gestalt Psychology

- European school of psychology founded in the early 1900s
  - Emphasizes our basic tendencies to actively organize what we see.
  - The mind forms a global whole with self-organizing tendencies.
  - When the human mind (perceptual system) forms a ***percept*** or ***gestalt***, the whole has a reality of its own, independent of the parts.

  - **"The whole is other than the sum of its parts."**
    - *The original famous phrase of Gestalt psychologist Kurt Koffka, "The whole is other than the sum of the parts" is often incorrectly translated as "The whole is greater than the sum of its parts" and thus used when explaining gestalt theory, and further incorrectly applied to systems theory.*
    - *Koffka did not like the translation. He firmly corrected students who replaced "other" by "greater". "This is not a principle of addition" he said. The whole has an independent existence.*

Content courtesy of Professor Mark Hale and Wikipedia

Organization in Visual Perception

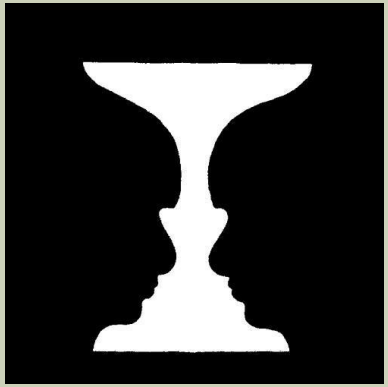
Our visual system is designed to impose organization on the complicated visual world.

- Gestalt Psychology
  - A fundamental principle in Gestalt Psychology is that humans have basic tendencies to organize what they see; without any effort, we see patterns, rather than random arrangements.
  - To make sense of the world we are constantly trying to separate and group things to give images meaning.
  - We take these processes for granted because they happen so naturally in normal situations.
  - They observed that properties emerge when two things are combined that were not found in any particular component.
    - Example: Think of a movie projector (filmstrip). There is no motion in the film but you perceive motion when it is presented to you at 24 frames per second!
    - They were the first psychologists to study how we visually organize sensory building blocks into meaningful units and patterns.

Content courtesy of Professor Mark Hale

Gestalt Psychology

- Gestalt psychologists noted that people always organize their visual field into figure and ground.
  - Figure: Things that stand out by nature of their size, intensity, color, movement, or perceived meaning.
  - Ground: All of the other items in the "background"
- All things being equal, we tend to see things that are lower in our visual field as figure and the upper part as ground.



Figure/Ground

- However, through selective attention, we are able to concentrate on some stimuli while ignoring others - thus changing our perception of the scene.

Content courtesy of Professor Mark Hale

Figure & Ground

- The explanation for figure-ground reversals has two components:
  1. The neurons in the visual cortex become adapted to one figure.
     - The alternative becomes the ground.
  2. People try to solve the visual paradox by alternating between the two reasonable solutions.

- We even perceive a figure-ground relationship when there is no clear-cut boundary!

- This category of visual illusions is known as illusory contours (a.k.a. subjective contours).

Theories of Visual Object Recognition

Theories of Visual Object Recognition

- We will concentrate on 3 theories of object recognition (many more exist):

  1. Template-Matching Theory
  2. Feature Analysis Theory
  3. The Recognition-by-Components-Theory
     - AKA Structural Theory
     - GEONS or GEON Theory

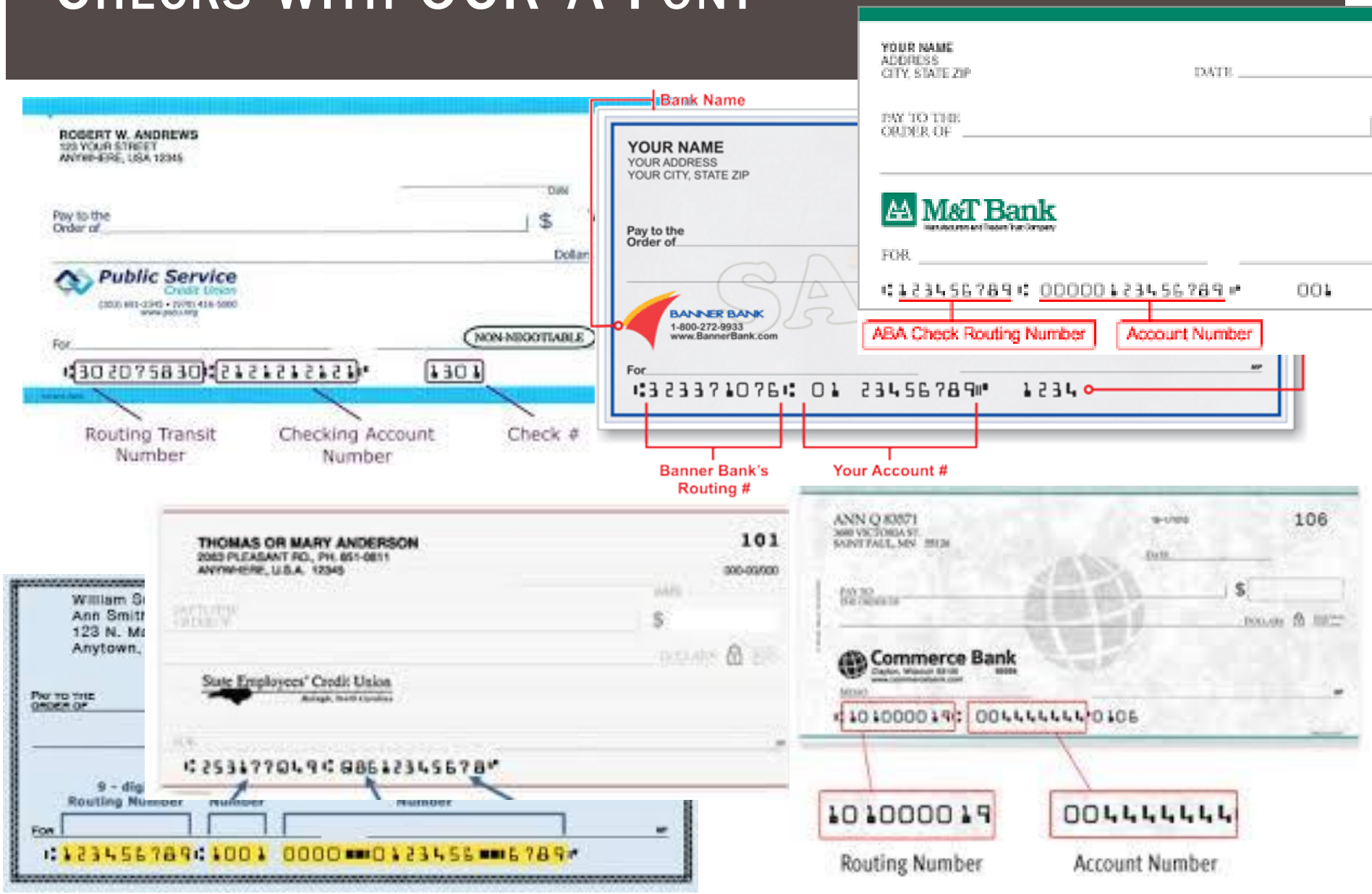
- Template-Matching Theory
  - When you see something you immediately compare it to with a set of "templates" in your head.
  - Example: Q will not fit into the O template because of the squiggly line at the bottom.

Content courtesy of Professor Mark Hale

Template-Matching Theory

- Many machine recognition systems are based on templates.
  - Numbers on a check. These numbers are specially designed to be recognized by check-sorting computers.
  - Each number has a constant, standardized shape that is distinctly different from the others, so the computer will not make a recognition error.
    - Back in the 1960s, a special font called OCR-A was developed that could be used on things like bank checks and so on. Every letter was exactly the same width (so this was an example of what's called a monospace font) and the strokes were carefully designed so each letter could easily be distinguished from all the others. Check-printers were designed so they all used that font, and OCR equipment was designed to recognize it too.
- Template matching works well with computer systems but falls short for explaining human object recognition.
- Problems with template-matching theory
  - It is extremely inflexible. If a letter differs from the appropriate template even slightly, the pattern cannot be recognized.

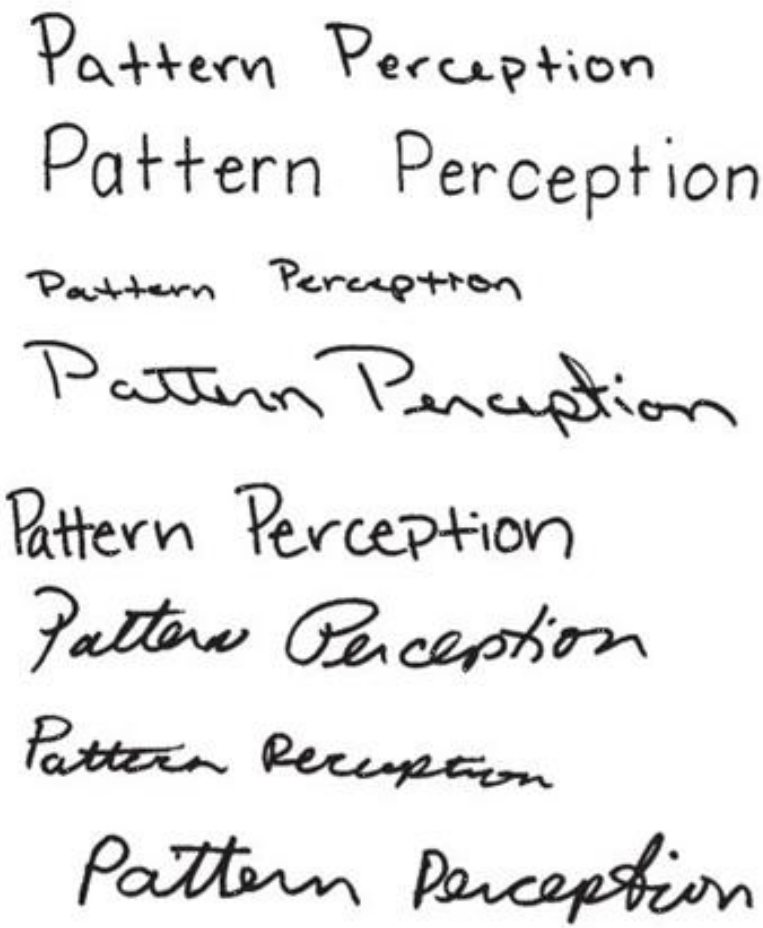
Checks with OCR-A Font



Human Recognition of Letters Cannot be Template-Based

- Consider all of these variations:

An example of variability in the shape of letters. Notice specifically the difference in the shape of the letter *P* in *Pattern*.



Content courtesy of Professor Mark Hale

Template-Matching Theory

- Problems with template-matching theory
  - Template models work only for isolated letters, numbers, and other simple 2D objects presented in their complete form.
  - Can't work with 3D models because of depth and viewing perspective!
  - Your perceptual processes must therefore employ a more flexible system than matching a pattern against a specific template.

Feature-Analysis Theory

- Feature-analysis theory (FAT) proposes a more flexible approach, in which a visual stimulus is composed of a small number of characteristics or components.
- FAT may explain the way we recognize letters of the alphabet.
  - This theory argues that we store a list of distinctive features for each letter.
  - For example, the letter R includes a curved component, a vertical line, and a diagonal line.
- When you look at a new letter, your visual system notes the presence/absence of the various features.
- It then compares this list with the features stored in memory for each letter of the alphabet.
- Features can vary slightly and still be recognized.

Content courtesy of Professor Mark Hale

A Feature-Analysis Approach

Eleanor Gibson proposed that letters differ from each other with respect to their distinctive features. The demonstration below includes an abbreviated version of a table she proposed. Notice that the table shows whether a letter of the alphabet contains any of the following features: four kinds of straight lines, a closed curve, an intersection of two lines, and symmetry. As you can see, the P and R share many features. However, W and O share only one feature. Compare the following pairs of letters to determine which distinctive features they share: A and B; E and F; X and Y; I and L.

Features	A	E	F	H	I	L	V	W	X	Y	Z	B	C	D	G	J	O	P	R	Q
Straight																				
horizontal	+	+	+	+		+					+				+					
vertical		+	+	+	+	+				+		+		+				+	+	
diagonal/	+						+	+	+	+	+									
diagonal\	+						+	+	+	+									+	+
Closed Curve												+		+			+	+	+	+
Intersection	+	+	+	+					+			+						+	+	+
Symmetry	+	+		+	+		+	+	+	+		+	+	+			+			

Source: Based on Gibson, 1969.

Feature-Analysis Theory

- FAT research mostly applies to letter and number recognition though it can help to explain simple designs and illustrations.

- Evidence for FAT?
  - Eleanor Gibson (1969) demonstrated that people require a relatively long time to decide whether one letter is different from a second letter when those two letters share a large number of critical features. (Psychological Evidence)
    - Example: Decision time when comparing two letters was longer when comparing P and R then P and Q.
  - Hubel and Weisel's experiment with cats found that certain neurons responded maximally to specific line orientations. (Neurological evidence)

- Therefore, we have specific neurons called feature detector cells that directly respond to certain features.

Content courtesy of Professor Mark Hale

Feature-Analysis Theory

- The Problems with FAT
  - FAT simply lists features contained in a stimulus regardless of their interrelationship.
    - - e.g., T and L have the same exact features!
  - FAT works with simple 2D letters and numbers but can't explain the complex shapes that occur in nature.
    - How could you recognize a horse according to FAT?
  - Movement further complicates this lack of ability to explain depth because it also distorts viewing perspective.

Content courtesy of Professor Mark Hale

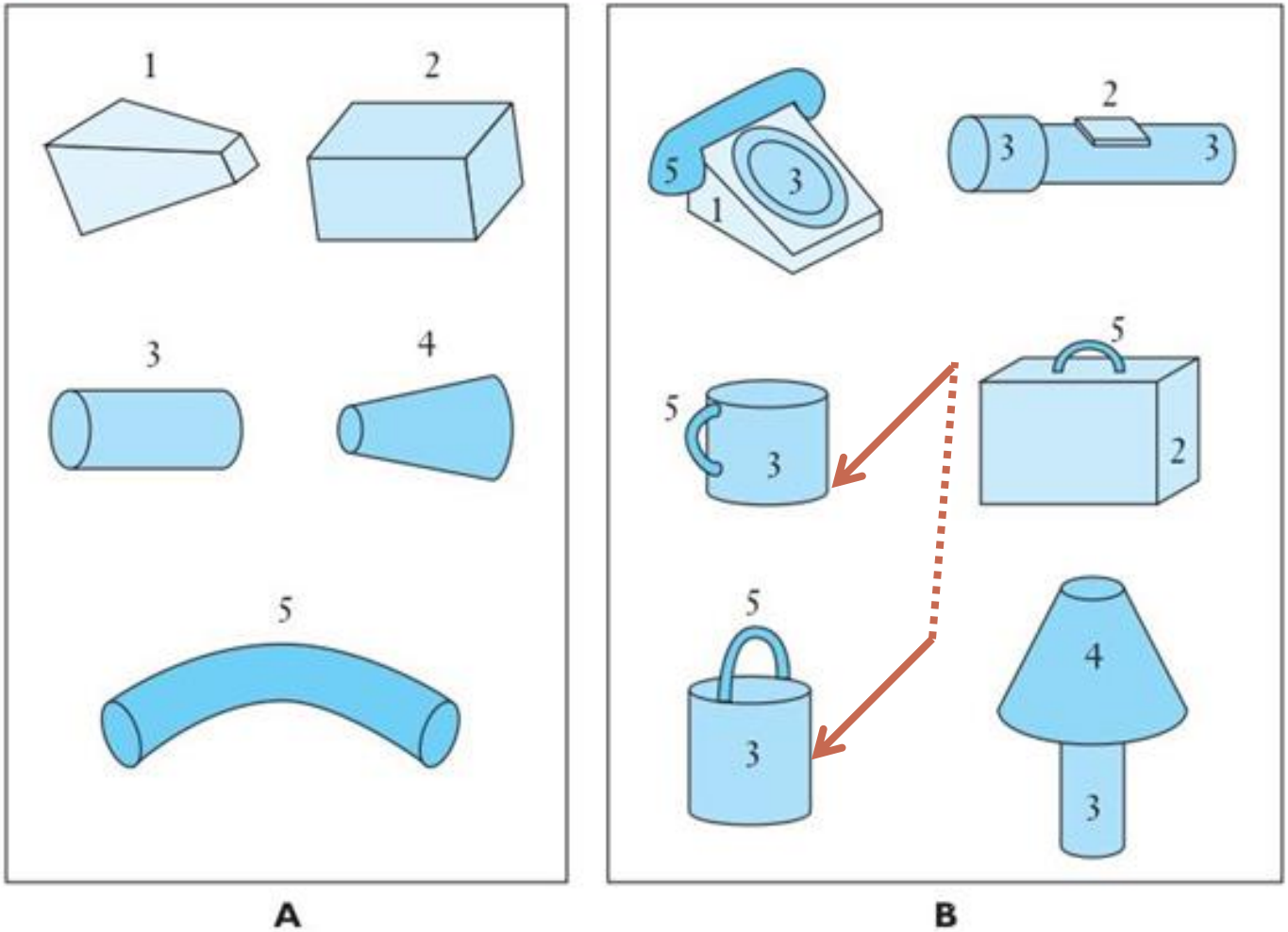
Recognition-by-Components Theory

- Recognition-by-Components Theory (RBC)
  - Also called Structural Theory or Geon Theory.
  - Irving Biederman and his colleagues have developed this theory to explain how humans recognize 3-D shapes.
- The basic assumption of RBC is that a specific view of an object can be represented as an arrangement of simple 3-D shapes called geons.
- Just as the letters of the alphabet can be combined into words, geons can be combined to form meaningful objects.

GEONS

FIGURE 2.5

Five of the Basic Geons (A) and Representative Objects that can be Constructed from the Geons (B).



Source: Biederman, 1990.

Cup is to pail in GEONS…

- as -

TOP is to POT in English.

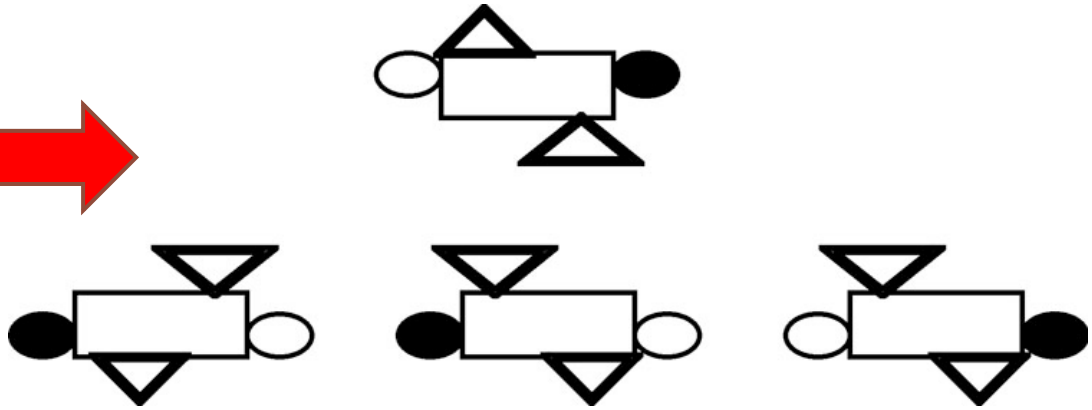
Recognition-by-Components Theory

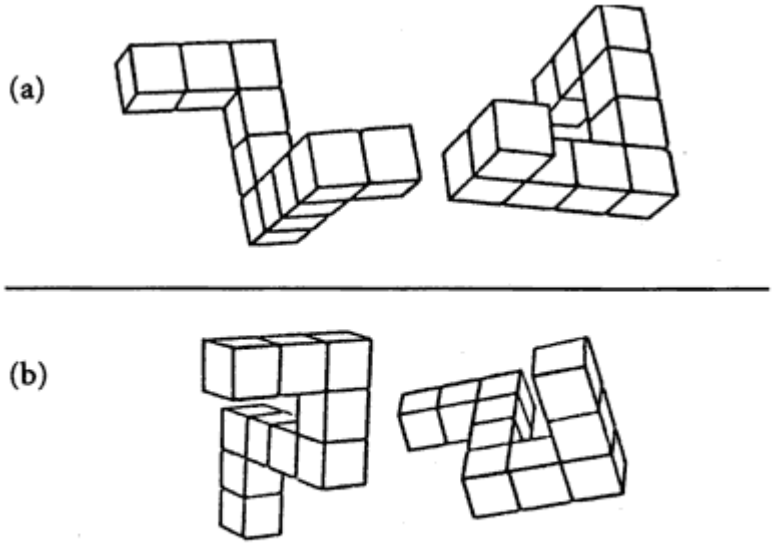
- In general, an arrangement of 3 geons gives people enough information to classify an object.
- RBC theory is similar to feature-analysis theory in that the component parts are used to identify the object as a whole.
- Remember though, RBC deals with the recognition of 3D objects as opposed to letters or numbers.

Content courtesy of Professor Mark Hale

Our Brain can Mentally Rotate Geons

Which one of the figures represents the original turned 180 degrees clockwise?





Mental Rotation Test—Are these two figures the same except for their orientation?

Top-Down vs. Bottom-Up Processing

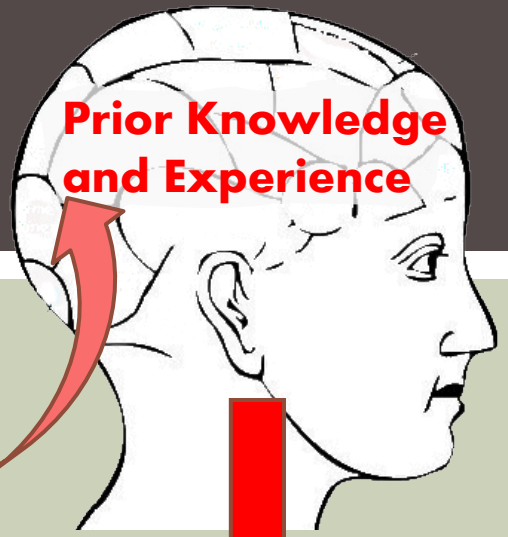
- Bottom-up processing emphasizes the importance of the stimulus in object recognition.
  - Specifically, the physical stimuli from the environment as registered on the sensory receptors.
  - The first part of visual processing may be bottom-up. However, an instant later top-down processing begins.

Top-Down vs. Bottom-Up Processing

- Top-down processing emphasizes how a person's higher-level mental processes influence/assist object recognition.
  - We expect certain shapes to be found in certain locations because of past experiences.
  - These expectations help us recognize objects very rapidly.
    - In that way, our previous knowledge and experiences work their way down to aid bottom up processing.
  - Top-down processing is especially strong when stimuli are incomplete, ambiguous, and when stimuli are registered for just a fraction of a second.
  - There is another type of processing worth mentioning here called anatomical (structural) coding.
  - Some researchers propose that specific structures along the route between the retina and the visual cortex may also play a role processing.
  - These structures may store information about the likelihood of seeing stimuli in a specific context.
  - We clearly need multiple types of processing to explain the complexities of object recognition.

Content courtesy of Professor Mark Hale

Top-Down Perception



Culture, Social Class

Occupation, Education

Values, Beliefs

Immediate Mental Set

Prejudices, Attitudes

Knoweldge, Vocabulary

Presence of Authority

etc.

Needs, Moods, Mental Health

Specific Life Experiences

Prior Stimuli Perceived

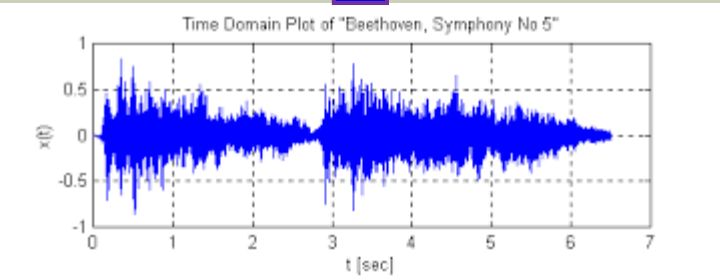
Personality, Temperment

Long-term Memory Schemas

Present Fatigue/Energy Level

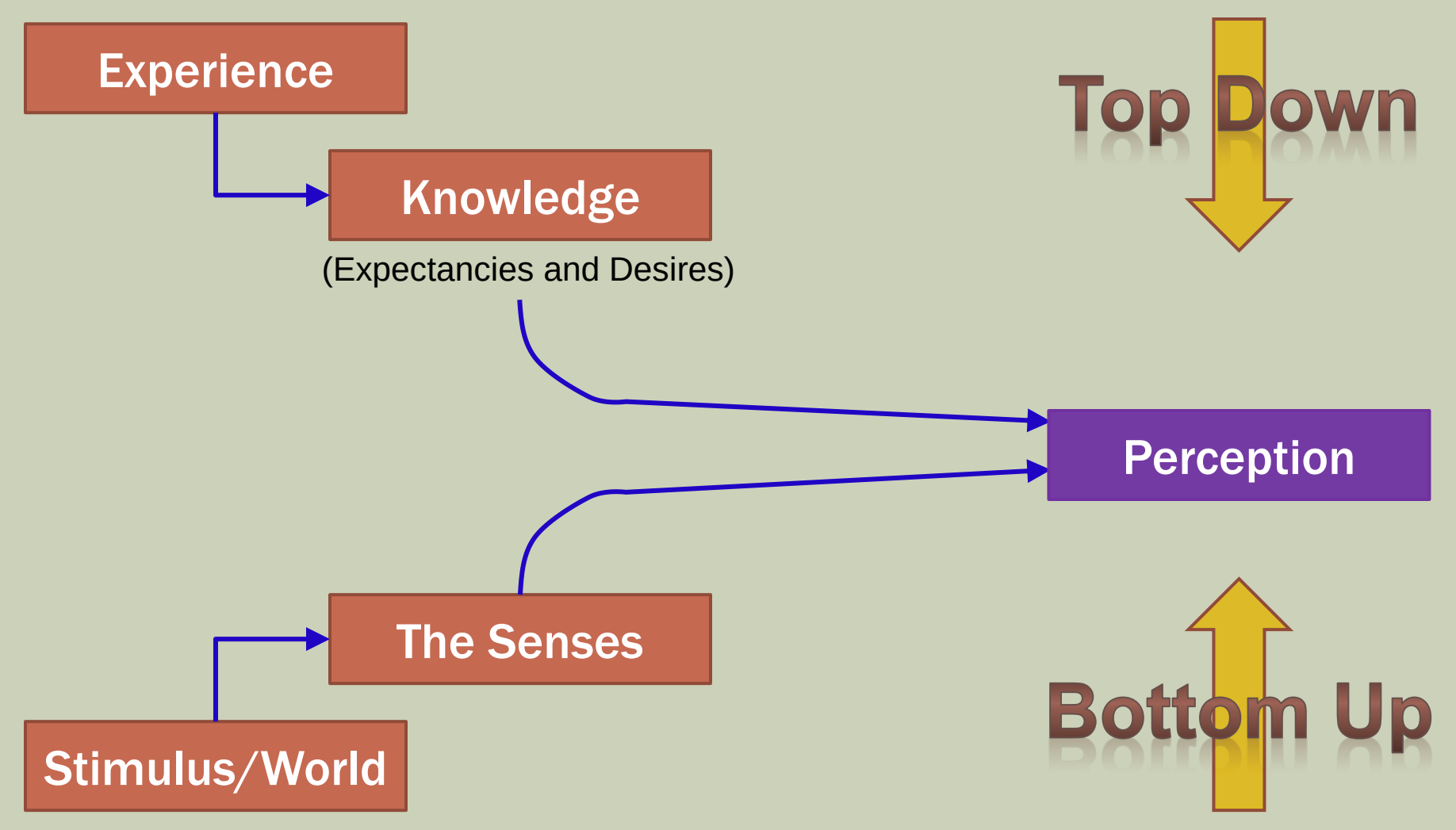
Perception

Stimuli Processing



Some content from Cardiff University Perception slides, http://slideplayer.com/slide/246517/

Another View of Top-Down Processing



From Wickens, An Introduction to Human Factors Engineering

Top-Down Processing & Reading

- Psychologists have realized for decades that a theory of recognition must include some factor(s) other than just the information in the stimulus. Why?

- If we were to examine the number of features in the average words, in the average sentences, over the average paragraph…
  - We would be analyzing ~5,000 features per minute. That just isn't possible.

- Furthermore, we can still read sentences when letters are wrong, swapped, or even missing entirely!

  - Example: "C4N U R3AD TH1S?"

  - Yes. However, research finds a very expected result… reading rate drops significantly!

Content courtesy of Professor Mark Hale

Face Perception

Face Perception

- One of most socially significant recognition tasks.
  - This is challenging because all faces share many commonalities (regardless of ethnicity) and there are an infinite number of faces to recognize!
- Recognizing faces vs. recognizing other objects
  - Many psychologists argue that most people perceive faces in a different fashion from other stimuli; in other words, face perception is "special."
- Evidence for the special nature of face perception
  - Young infants track the movement of a photographed human face more than other similar stimuli.
  - People are significantly more accurate in recognizing facial features when they appear in the context of a whole face, rather than in isolation.
  - In contrast, when they judged houses, they were just as accurate in recognizing an isolated house feature (such as a window) as in recognizing house features within the context of a complete house.
  - Faces apparently have a special, privileged status in our perceptual system.
  - We recognize faces on a holistic basis— that is, in terms of their overall shape and structure.

Content courtesy of Professor Mark Hale

"Face Module"

- A face module makes sense from an evolutionary perspective. It helps babies distinguish “friend from foe” and perhaps a mother from a stranger.

- Neuroscience Evidence on Face Recognition
  - Researchers have reported that certain cells in the inferotemporal cortex of monkeys respond especially vigorously to a photo of another monkey’s face.
  - fMRI studies show that the brain responds more quickly to faces presented in a normal (upright) position than to faces presented upside-down.

- How would computers perform face recognition?

- Since this face module is biologically hard-wired, it becomes a harder task for computer to emulate.

Content courtesy of Professor Mark Hale

Speech Perception

Speech Perception

- Take a moment to think about the truly complex nature of recognizing human speech.
  - During speech perception, your auditory system must record the sound vibrations generated by someone talking; then translate these vibrations into a sequence of sounds that you perceive to be speech.
  - The average adult who speaks English produce about 15 sounds every second (~900 sounds per minute).
  - You must take these sounds and compare them against the thousands of words you have in your lexicon.
  - You must also distinguish between signal and noise!
- Characteristics of Speech Perception
  - Phoneme - the most basic unit of speech sound.
  - Phonemes combine to make morphemes, words, and sentences. There are 45 phonemes in the English language.
- Important characteristics of speech perception
  - Word boundaries
  - Pronunciation variability
  - Context
  - Visual Cues

Content courtesy of Professor Mark Hale

Speech Perception

- Word Boundaries
  - Listeners can impose boundaries between words, even when these words are not separated by silence.
  - Have you ever noticed when you listen to conversations in an unfamiliar language it appears to more continuous than that of your native language?
  - In most cases, however, the actual acoustical stimulus of spoken language shows no clear-cut pauses to mark the boundaries!
    - An actual physical event — such as a pause — marks a word boundary less than 40% of the time.
  - It is our knowledge of the words that allow us to perceive these word boundaries.

top-down!

- Variability in Phoneme Pronunciation
  - Phoneme pronunciation varies tremendously! Qualities in our voices (pitch and tone) as well as production rate vary greatly!
  - We also fail to produce these phonemes precisely (systematic vs. accidental)
  - The phoneme that you produce will always vary due to a phenomenon known as coarticulation.
  - Coarticulation is the fact that your mouth will remain somewhat the same shape as it was in pronouncing the previous phoneme.
  - Furthermore, your mouth is also preparing to pronounce the next phoneme at the same time.
    - For example, notice that the d in idle sounds different from the d in don't.

Content courtesy of Professor Mark Hale

Speech Perception

- Context and Speech Perception
  - Context allows listeners to fill in missing sounds.
  - Just as we saw with visual perception, top-down processing plays a huge role in language perception

- Researchers showed that people are skilled at using the meaning of a sentence to select the correct word from several options.
  - A cough sound was inserted to cover up a phoneme. ← This is known as "masking"

1. It was found that the <cough>eel was on the axle. ← Subjects heard "wheel"
2. It was found that the <cough>eel was on the shoe. ← Subjects heard "heel"
3. It was found that the <cough>eel was on the orange. ← Subjects heard "peel"

- Phonemic restoration is a well documented kind of illusion.

Content courtesy of Professor Mark Hale

Visual Cues and Speech Perception

- Visual cues from the speaker's lips and face help us to resolve ambiguities from the speech signal, much as linguistic cues help us choose between wheel and heel.
- Think about those who are adept in lip reading!
- Researchers showed participants a video of a woman whose lips were producing simple syllables, such as "gag." Meanwhile, the researchers presented different auditory information, such as "bab."
  - When the observers were asked to report what they perceived, their responses usually reflected a compromise between these two discrepant sources of information.
  - In this case, the listeners typically reported hearing the word "dad."
- The McGurk Effect refers to the influence of visual information on speech perception, when individuals integrate both visual and auditory information.
- Therefore, McGurk suggests that speech perception is a multimodal.
- Think how much easier it is to understand people in person than on the telephone!

Content courtesy of Professor Mark Hale

Speech Perception Theories

- Most current theoretical approaches to speech perception fall into one of two categories.
  - The Special Mechanism approach
  - The General Mechanism approach

- The Special Mechanism approach
  - Also known as the "Speech is special" approach
  - Humans are born with a specialized device that allows us to decode speech stimuli.
  - Therefore, we must process speech sounds more quickly and accurately than other auditory stimuli, such as instrumental music.
  - This phonetic module would presumably enable listeners to perceive ambiguous phonemes accurately.
  - It would also help listeners to segment the blurred stream of auditory information that reaches their ears, so that they can perceive distinct phonemes and words.
  - This approach argues that the brain in organized such that speech perception does not rely on other general cognitive functions.

Speech Perception Theories

- The General Mechanism Approaches
  - Although some still favor the special mechanism approach, most theorists now favor one of the general mechanism approaches.
  - This phonetic module would presumably enable listeners to perceive ambiguous phonemes accurately.
  - It would also help listeners to segment the blurred stream of auditory information that reaches their ears, so that they can perceive distinct phonemes and words.
  - This approach argues that the brain is organized such that speech perception does not rely on other general cognitive functions.
  - The general mechanism approaches argue that we can explain speech perception without proposing any special module.
    - In other words – no difference in the processing of speech vs. non-speech sound.
  - Evidence supporting general mechanism approaches
    - Using event-related potentials (ERPs), adults show the same sequence of shifts in the brain's electrical potential, whether they are listening to speech or to music.
    - People's judgments about phonemes are definitely influenced by visual cues.
    - However, if we have this special module why would we need (or benefit) from visual cues?

Content courtesy of Professor Mark Hale

Speech Perception Theories: Conclusion

- The general mechanism approaches argue that we can explain speech perception without proposing any special module.
  - In other words – no difference in the processing of speech vs. non-speech sound.

- Evidence supporting general mechanism approaches
  - Using event-related potentials (ERPs), adults show the same sequence of shifts in the brain's electrical potential, whether they are listening to speech or to music.
  - People's judgments about phonemes are definitely influenced by visual cues.
  - However, if we have this special module then why would we need (or benefit) from visual cues?

Content courtesy of Professor Mark Hale

What About Background Noise?

- How do you recognize speech on a phone call with background noise?
- How can Shazam focus on the audio signal of a song while filtering out background noise?

(for a very interesting discussion, see http://coding-geek.com/how-shazam-works/ – August 2015)

- Turning up the volume will not work, because the volume of the noise is amplified as well
- The strategy?
  - Isolate the signal from the noise and amplify the signal only
  - If you are sending an auditory signal, realize that low frequency noise masks high frequency signals better than the reverse. Hence low frequency transmissions are less maskable. *(for example, see https://www.youtube.com/watch?v=Yfx69eTiVSc)*

More Auditory Interactions – Alarms

Effective Alarms

- Often computer systems need to transmit information via auditory alarms
  - Better than an alarm message as it is easier to "close your eyes" than to "close your ears"
  - Nuclear power plant management systems
  - Air traffic control systems
  - Intrusion monitoring systems

- Criteria for Alarm Design
  - Alarms must be heard over background, ambient noise
  - The alarm should not be above the danger level for hearing, i.e., 85 to 90 decibels
  - The alarm should not be overly startling or abrupt
  - The alarm should not disrupt the perceptual understanding of other signals.
    - "I was flying [an airplane] at night when my peaceful revelry was disturbed by the stall audio warning, the [control] stick shaker, and several warning lights. The effect was exactly what was ***not*** intended. I was frightened numb for several seconds and drawn off instruments trying to work out how to cancel the audio/visual assault, rather than taking what should have been instinctual actions." – British pilot
  - The alarm should be informative, signaling to the listener the nature of the emergency and, ideally, corrective actions

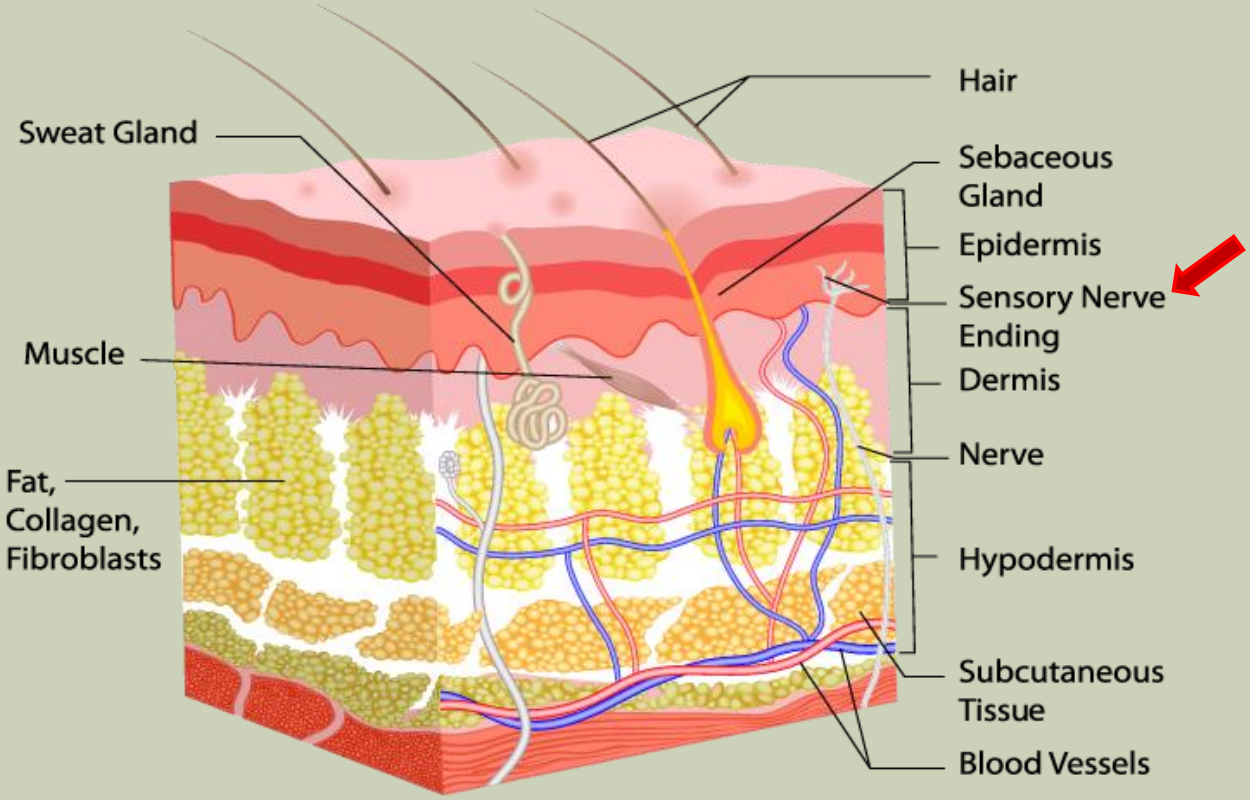
From Wickens, An Introduction to Human Factors Engineering

Touch

Haptic Senses

- Lying just under the skin are sensory receptors that respond to pressure on the skin and relay their information to the brain.
- The receptors respond to the subtle changes in force applied by the hands and fingers (or other body parts) as they interact with physical objects in the environment.
- Haptic: relating to or based on the sense of touch.
- Two types of Haptic feedback.
  - **Tactile**
  - **Kinesthetic**
  - Together, they provide information about object qualities, bodily movements, and their interrelationships.

TACTILE



- **Tactile**: subjectively, the things you feel in your 'fingers' etc., or on the surface. Tactile sensing, for example, is needed to determine the local shape and texture of objects.

- The tissue (for example in your fingers), has a number of different sensors embedded in the skin and right underneath it. They allow your brain to feel things such as vibration, pressure, touch, texture etc.

Image source: http://londondermatology.com/wp-content/uploads/2014/04/what_does_skin_do.png

KINESTHETIC

- **Kinesthetic**: Deals with the sense of body movements and muscle feelings. The things you feel from sensors in your muscles, joints, tendons. Weight, stretch, joint angles of your arm, hand, wrist, fingers, etc. Imagine holding a coffee-mug in your hand.

- Kinesthetic feedback tells your brain the approximate size of the mug, it's weight, and how you are holding it relative to your body. (Imagine holding a mug. Is not muscle feedback in play here?)

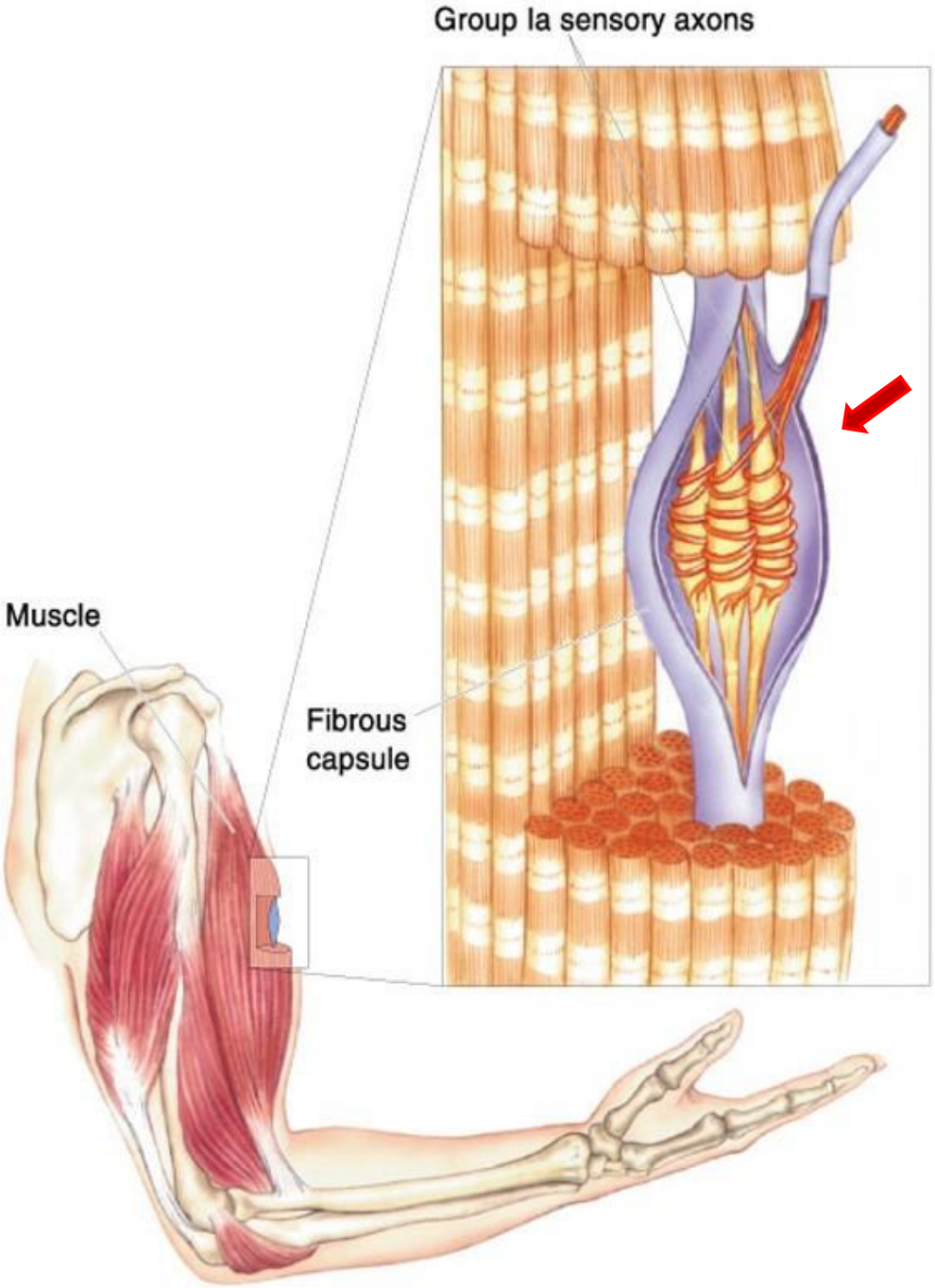


Image source: https://classconnection.s3.amazonaws.com/979/flashcards/4502979/jpg/proprioceptors-1455DE7100D65419042.jpg

Tactile Considerations

- The tactile "feel" of keyboards. Touch screen keys versus mechanical keys

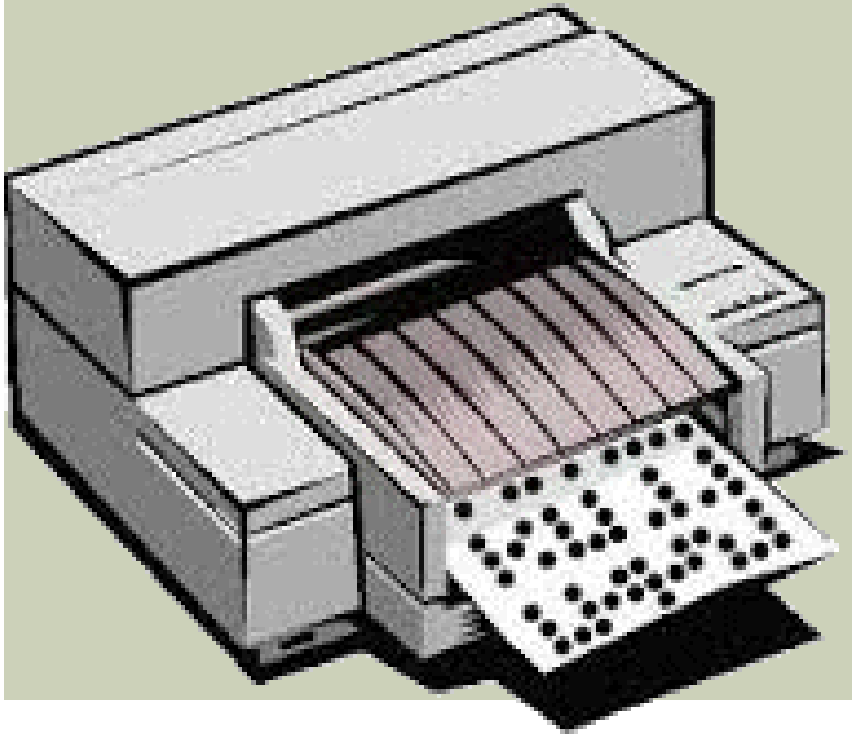
Tactile Considerations

- Gloves designed with tactile sensitivity

Tactile Considerations

- Braille for the visually impaired

Keyboards

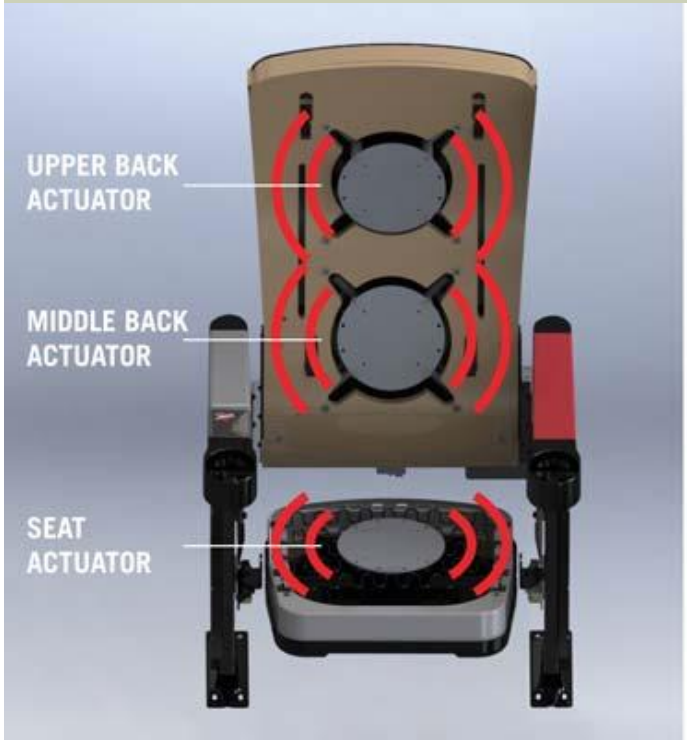


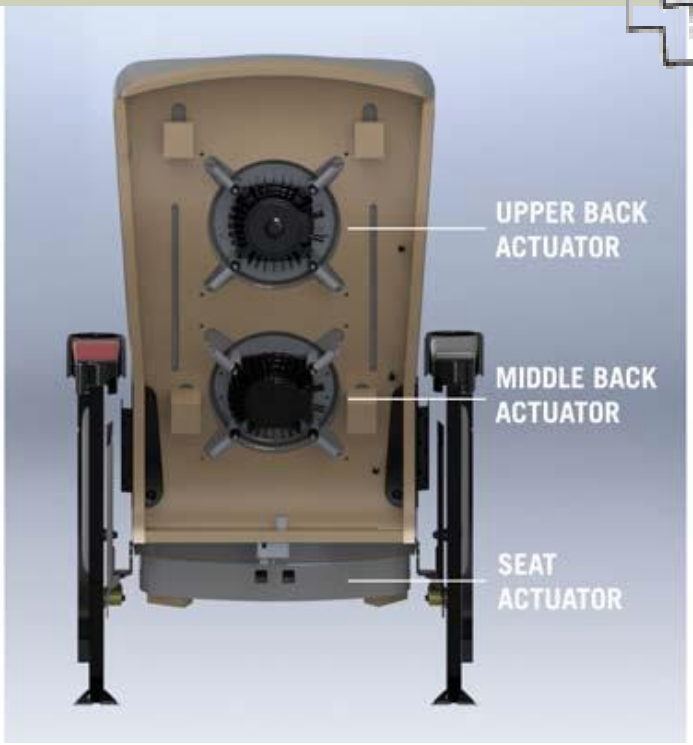
Printers

Tactile Considerations

- Vibrations as a form of feedback

D-Box Movie Theatre Seats





Vibrating Phones