Visual Perception

The Visual System

The human visual system can be divided into two stages:

- Physical reception of light
- Processing and interpretation

The human visual system has both strengths and weaknesses:

- Certain things cannot be seen even when present
- Processing allows images to be constructed from incomplete information

Light passes through the *cornea* and is focussed by the *lens*, producing an inverted image on the *retina*.

The *iris* regulates the amount of light entering the eye.

The retina is covered with *photoreceptors*. These are of two types:

rods	High sensitivity to light Monochrome Low resolution
cones	Limited sensitivity to light Colour (red, green blue) High resolution

The eye contains:

- around 6 million cones, most of which are situated within the fovea.
- around 120 million rods, most of which are situated around the *periphery* of the retina.

The photo-receptors are all connected to the *optic nerve*, which carries visual information to the brain.

There are no photo-receptors in the area of the retina around the optic nerve.

Thus there is a *blind-spot* at this point.

We are not usually aware of the blind spot because our brains 'fill in' the missing part of the image.

The Visual Field

When we look at a scene:

- we view a very small area using *foveal* vision
    - high resolution
    - colour
- the remainder of the visual field is seen with *para-foveal* or *peripheral* vision
    - lower resolution
    - little or no colour perception

Because we can only focus on a very small area at a time, we *scan* scenes, i.e.:

- move our eyes rapidly, focussing on different parts of the scene in turn
- build up a composite image over time

We have a good *spatial memory* which allows us to remember (e.g.) where a particular piece of information is located on a page or screen.

Recalling the *location* of a piece of information appears to use less short-term memory than recalling the information itself.

Hence the page or screen acts as a kind of *external memory*.

Brightness Perception

Luminance is a physical property that can be measured.

The luminance of an object depends on:

- The amount of light falling on to its surface
- The reflective properties of the surface(s).

Contrast is related to luminance. It is the difference in luminance between the brightest and darkest areas of an image.

Perception of brightness is subjective.

The human visual system compensates for bright or dark conditions by:

- varying the iris aperture to regulate the amount of light reaching the retina
- using the lens to direct the image onto different parts of the retina:
    - in bright conditions, light is focussed on the fovea, giving high resolution and colour vision.
    - in dark conditions, focus is shifted onto the periphery, giving greater sensitivity but reducing resolution and colour perception.

Thus it is difficult to measure/quantify human perception of changes in brightness.

However, we can measure the *just noticeable difference* (JND) under various conditions.

Increasing the brightness:

- improves visual acuity, but also...
- increases perception of flicker - flicker may become obvious even at higher frequencies.

Colour Perception

Light is an electromagnetic wave, like radio waves but of much higher frequency.

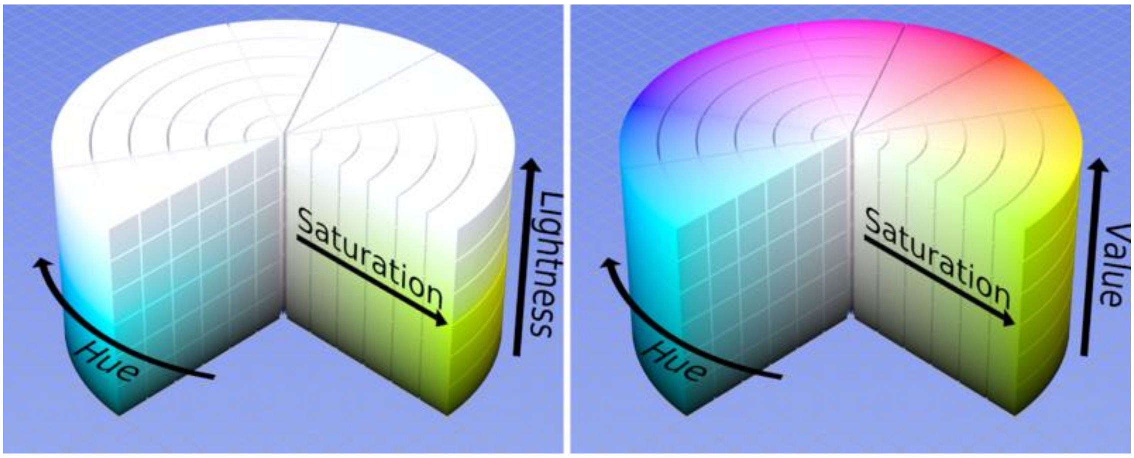
The visible spectrum ranges from violet at the high-frequency (short wavelength) end to red at the low-frequency (long wavelength) end.

The human eye can distinguish about 150 *hues* within the visible light spectrum.

However, the total number of *colours* we can distinguish is much higher.

This is because:

- Each of the pure hues can be mixed with white in various quantities to produce other colours.
    - For example, if pure red is mixed with white, the result is a shade of pink.
    - We refer to the spectral hues as *saturated* colours.
    - When mixed with white, we refer to them as *partially-saturated* or *de-saturated* colours.
- The *brightness* of each hue can be varied.
    - For example, if we progressively reduce the brightness of a shade of red, we obtain darker reds and eventually black.



Taking into account all three factors - *Hue*, *Saturation* and *Brightness* - we are able to distinguish around 7 million colours.

Note that:

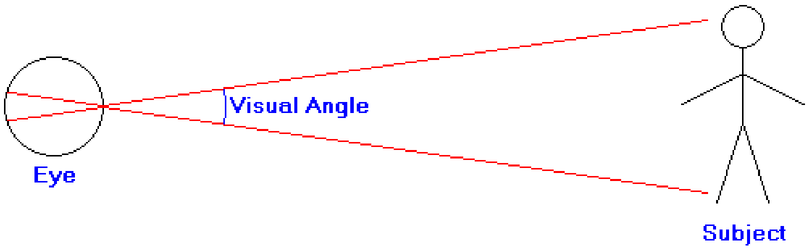
- Colour sensitivity is greatest in the fovea, where cones predominate.
- Only around 2-3% of the cones are sensitive to blue light.
    - Therefore, discrimination between different shades of a colour is worse for blue than for the other colours.
- The eye is most sensitive to green light.
- Around 8% of males and 1% of females have some form of colour-blindness (usually red-green colour-blindness).
- Colour perception is normally *pre-attentive* - humans with normal colour vision perceive and discriminate colours immediately and without conscious effort.

Distance Perception

A small object that is close by may have the same visual angle as a larger object that is further away.

However, we are good at gauging the size of objects, even when we see them at extremes of range.

Clearly, visual angle alone does not determine perceived size.



Factors affecting our judgement of size include:

- *Binocular vision* - the difference in the image seen by each eye can be analysed to gauge distances
- *Head movement* - small changes in viewing position produce changes in view that allow distance to be gauged
- *Monocular Cues*:

Relative size	Relative height
Relative clarity	Interposition / Occlusion
Texture gradient	Linear perspective
Shadow cues	

- *Familiarity / Pattern-Matching*

Binocular Vision

When viewing a scene, each eye has a slightly different view.

This is known as *retinal disparity*.

The closer an object is, the greater the retinal disparity.

In order to view objects at different distances, we adjust our focus.

This is known as *convergence*.

Feedback from the eye muscles tells us how much adjustment has been made, and hence how far away the object is.

These two factors - retinal disparity and convergence - together provide us with a great deal of information about depth and size.

Head Movement

The information obtained from retinal disparity and convergence can be enhanced by moving the head.

Feedback from our muscles tells us how far we have moved, and hence the distance between two viewing positions.

By comparing retinal images and eye-convergence information from two (or more) positions, we can obtain greater information about depth and size.

Relative size

Where an image features several objects of similar shape the tendency is to assume that the smaller objects are further away.

Relative height (height in field)

- Where the base of a shape is higher than that of a similar shape, the one with the higher base is assumed to be further away
- This cue relates objects to the horizon

Relative clarity

We cannot see distant objects with as much clarity as near-by objects, so reduced clarity in some parts of an image implies distance

Interposition (occlusion, superimposition, overlay)

When one object is interpreted as obscuring part of another one, the one which seems to be obscured is seen as being further away.

Texture gradient (or detail perspective)

Texture gradient can be seen as a combination of linear perspective and relative size

Areas in which particular shapes are densely packed will appear more distant than areas in which similar shapes are less densely packed.

Perspective

We find it hard to avoid interpreting converging lines as indicating linear perspective.

Depth Cue Shadow

In processing visual information, we tend to assume:

- Light comes from above
- Objects are viewed from above and not from below

Familiarity / Pattern-Matching

When processing visual information, we are primarily seeking to *identify* objects in the visual field.

This also applies to judgements of distance and size.

We expect certain objects to be of a particular size, and use this information to help us interpret a scene.

Pre & Post Attentiveness

Eye-movement and focus typically take around 200ms.

Visual tasks that involve eye-movement and focus are said to be *post-attentive*.

Some visual tasks don't require transfer of focus and can be accomplished immediately. They are said to be *pre-attentive*.

- Pre-attentive:
    - simple colour-identification tasks
    - e.g., finding an object of a particular colour among a group of coloured objects
- Post-attentive:
    - identifying/comparing size, shape, etc.
    - task involving a combination of factors, e.g., colour plus size

Reading

Research shows that people:

- read from a computer screen around 25% more slowly than from printed material.
    - adults typically read print at around 250 words per minute.
- 'scan' material on screen more than they do printed material.
    - e.g., read progressively less of each line/paragraph as they move down the screen
- dislike scrolling
- dislike 'wordy' text

This may be because computer screens, and the typography and layout often used on computers, do not adequately support adult reading strategies.

When children learn to read, they initially read *linearly*, i.e.:

- start at the beginning of the sentence
- read each word in turn
- identify the meaning of each word (letter-by-letter if necessary)
- identify the meaning of the sentence

This is a very slow and inefficient method of reading.

As children become more proficient at reading, they learn to *scan* text by *spotting key-words*.

This process involves a number of stages:

Identify a word or character	Studies show that long words are recognised as quickly as single characters. This suggests that words are recognised by shape, not by identification of characters.
Guess the meaning of the phrase or sentence	
Confirm/disprove the guess	The reader jumps forward through the text, looking for words or characters that will confirm/disprove the guess. Forward jumps are known as *saccades*.
Revise the guess if necessary	If the guess cannot be confirmed, it may be necessary to back-track and revise the guess. Backward jumps are known as *regressions*.

Factors that affect the readability of text include:

Font-style and capitalisation	Pattern-recognition is crucial to reading, so type-faces with distinct patterns are easier to read than others. Block capitals are particularly hard to read.

Font size	Font sizes from 9-12 point are equally legible (assuming proportional spacing); larger and smaller sizes are less legible.
Character spacing	Proportionally-spaced text is easier to read than text with fixed-spacing.
Line length	Lengths of between 2.3" (58mm) and 5.2" (132mm) are equally legible.
Contrast / Luminance	Black text on a white background is easier to read than (e.g.) white text on a black background.

Morkes and Nielsen (1997) asked subjects to rate several versions of a web-page for usability.

They used an existing tour web-site and created several versions which contained the same information but presented/worded differently.

The ratings given to the various versions (compared with the original) were as follows:

abbreviated text	rated 58% better
text split into single lines	rated 47% better
objective language only	rated 27% better

A version of the page that combined all three approaches was rated 124% better than the original.

Implications for Interface Design

2D versus 3D:

It has long been assumed that 3D displays will eventually replace 2D displays in most applications.

However, 2D displays continue to be used for most purposes.

One reason may be the cost and availability of 3D technology compared with 2D.

However, research and practical experience suggest that, for many applications, 2D remains a more appropriate choice than 3D.

- Cognitive/Perceptual Effort:
  - In 3D, some cues may be absent, incomplete, or misleading, e.g., eye-convergence
  - Presenting incomplete/incorrect 3D information increases the cognitive load.
- Abstraction versus Realism:
  - We are used to seeing abstract 2D views, but expect 3D views to be realistic.
  - 2D views are better suited to scaling, zooming, etc.

3D is suited to applications that involve the visualisation of real objects and their orientation/relationship to other objects, e.g.,

- medical imaging: the shape and position of an organ is easier to understand from a 3D model than from a 2D X-ray

- mechanical modelling, allowing engineers to view a component and consider its fit/interaction with other components
- architectural modelling, allowing designers to view the shape of a structure and examine sightlines, layout of internal spaces, etc.

2D is suited to applications that involve exploration or visualisation of abstract data, e.g.,

- a computer's file-structure
- Financial and commercial data

3D displays are increasingly used in visualisations of complex, multi-dimensional data.

However, many designers recommend avoiding 3D if alternative 2D representations can be used.

Screen Usage and External Memory

By presenting lots of information at once we can use the screen as external memory and reduce the load on the user's short-term memory.

However, studies suggest that users 'skim' material on web-pages, and that it is usually best to minimise the amount of material per page.

Be clear how the material on screen is to be used:

- For tasks that involve comparison, analysis, decision-making, etc., it is usually best to place as much relevant information as possible on screen
    - the screen serves as external memory
- For most other purposes, bear in mind that users won't read lengthy passages from a screen so minimise the amount of text and make it as readable as possible.

When presenting complex data:

- Colour is a good choice of cue because it is perceived pre-attentively, but...
    - some people are colour-blind, so provide redundancy and/or allow customisation
- Shape and size are perceived post-attentively, but can still be used effectively
- Brightness is a poor cue because humans adapt to it

When presenting text:

- minimise the amount of text by removing all unnecessary material
- avoid descriptions, etc.
- break-up the text using (e.g.) bullet-points

If using longer passages of text is unavoidable, try to minimise its complexity.

There are a number of methods which can be used to measure the complexity of a passage of text

- Average reading time
    - A group of people are asked to read the text, and the average time taken is noted.

- Cloze Technique
    - Subjects are asked to read a piece of text in which every fifth word is blanked out.
    - The index is based on the percentage of blanked words that are guessed correctly.
    - Texts with simple, predictable structures usually obtain high scores.
- Fog Index, Flesch Reading Ease Test, etc.
    - Take into account factors such as average length of words and sentences, percentage of 'complex' words, etc.
    - Grade texts in various ways, e.g.
        - On a scale from 6-17, indicating the age at which pupils should be able to read text of that complexity.
        - By US school grade, e.g., a score of 8 indicates a text an 8th-grader should be able to read.
        - On a 100-point scale

Tests such as the Fog Index generally give useful measures of complexity for US English, but arguably not for other languages and dialects.

However, they have the advantage they allow texts to be analysed quickly and easily using automated tools:

- Many word-processors have built-in readability tools, e.g., MS Word includes a Flesch Reading Ease tester
- Tools are available online, e.g. `http://gunning-fog-index.com/`