SE3VR11 Virtual Reality
Prof Richard Mitchell

Perception and Technology

In Virtual Reality we create an imaginary world to inhabit

- Important to understand how humans perceive 'worlds'
- Covers Visual, Auditory and Haptic senses

The notes for this course have been developed over many years by many people. My thanks to

Max Parfitt, Faustina Hwang, Andrew Dunk, and Adrian Haffegee

Introduction

PMS has defined VR, introduced some terminology, covered some of the history and applications

He has listed the 4 key elements (Sherman & Craig, 2003)

- A Virtual Environment (world)
- Immersion
- Sensory feedback
- Interactivity

And the Components of a VR System.

- Input and Output Devices
- VR Engines
- Databases

Virtual Reality and Games

Many are interested in computer games …

Often these games portray a world visually and auditory

They have controllers which allow exploration of the world and interaction with the world and others there

To an extent they are Virtual Reality, but

- A flight simulator game at home is ok,
- But not the same as using a real flight simulator
  - (with hydraulically operated mockup of a cockpit)

But Headsets are increasingly available for video games

Games driving VR

In 2016, $5.1billion will be spent on VR gaming

http://fortune.com/2016/01/05/virtual-reality-game-industry-to-generate-billions/

Not surprising therefore that the games industry has helped considerably re VR development

As such, for the individual part of the coursework, you will all use the Unity software https://unity3d.com/5

Good for making simple worlds, with inbuilt physical modelling, interaction, movement and exploration

For more detailed models, other software is better

So in Groups make more complex worlds – individuals in groups can focus on other packages.

Focus of My Lectures

I shall focus on how a VR world can be produced

Note Technology provides only a model of the world

Sufficient to 'fool' our senses

This requires an understanding of how we perceive

Hence I look at the senses first then the technology

First, however, how can we quantify VR?

This relates to some of the four key elements..

Then I justify why we need to know of 'Human Factors'

Quantifying VR

What makes one VR experience better than another?

Graphical realism? Frame rate? A good storyline?

How can you describe what's good or bad about VR to someone?

Common terms are: Presence and Immersion.

Used to describe physiological and psychological integration of a subject when using a virtual reality system.

But…

Immersion and Presence

As loosely defined as "Virtual Reality"

Most common definitions are:

- Immersion refers to technological sophistication
- e.g.: 'a measure of … the [VR] system's capacity to reproduce the physical sensation of the real world'

(S. Bouchard 2009)

- Presence refers to a mental state
- e.g.: 'the physical feeling of being in a place other than the physical location that the user finds him/herself in.'

(T. Sheridan 1996)

Determinants of Immersion

How well the VR technology delivers aspects of the VE to the user; measurable and comparable between systems

Determinant of Immersion	Description
Sensory Channels	No. different kinds of sensory stimuli produced by the system
Field of Regard	Directions in 3D from which a stimulus can originate (sound could come from any direction; touch could occur anywhere on body)
Fidelity	The quality of the displays, resolution etc.
Correspondance	How closely user's movements match corresponding environment changes, view position, hand movements etc

Determinants of Presence

How well user feels as though they are a participant within the environment

Determinant of Presence	Description
Interaction	The ease and extent of available interaction
Duration	The length of exposure to a simulation
Self representation	The user's ability to see their own body and actions
Consistency	The uniformity of sensory information. The position something is seen should match the position it is felt
Behaviour	The realism of objects' behaviour in the environment

Distracters and Breaks in Presence

What reduces presence and immersion …

Distracters are stimuli which seem out of place and remind the user they are not really part of the simulation

A 'Break in Presence' according to Mel Slater* occurs when the user stops responding to virtual stimuli and starts noticing real stimuli.

Examples:

Screen flicker, unrealistic/unexpected VE behaviour, sickness, loud noise etc.

(Interesting psychological research on haptics in VR)

*M. Slater, "Presence and the Sixth Sense,," Presence: Teleoperators & Virtual Environments, vol. 11, pp. 435-440, 2002.

Faithfulness and Believability

An alternative way of classifying VR experience.

More appropriate for describing design of VR systems

Faithfulness:

- Refers to how accurately a simulator recreates real world stimuli.
- Similar to the definition of immersion as technological sophistication

Believability:

- Refers to user experience and psychological state. A Believable VR simulator is not necessarily "faithful"
- Similar to most definitions of Presence.

Why Study Human Factors?

"To adequately create the illusion of reality we must understand both the sensory process and the inference process"

- S. Aukstakalnis

"Designing or selecting a graphics display cannot be done meaningfully without first understanding the human visual system. An effective graphics display needs to match its image characteristics to those of the user's ability to view the synthetic scene."

– Burdea & Coiffet

Why Study Human Factors?

i.e. Virtual Reality is about working out where you can cheat.

To create a believable Virtual Reality you must first understand how the senses perceive reality.

This sets minimum technological requirements

Also sets upper thresholds which need not be exceeded

We will look at what actual happens

Then consider the technology that is suitable

Sensory Perception – Some Jargon

Cue Theory:

A theory [there are others] of how complex concepts are inferred from sensory stimuli

In summary, cue theory states: 'Humans learn to make connections between key events in the sensory information streams (cues) and real world phenomena'

You will hear VR people using the term 'cues' a lot

Useful way of thinking about what form feedback should take in VR

More Jargon

Primary and Secondary Qualities:

Primary Qualities can be perceived by many senses.
Position is visual, audible and tactile

Secondary Qualities are unique to a single sense
Red is unique to vision as bitterness is unique to taste

Primary qualities can be more difficult as different sensory displays must agree or conflict can arise

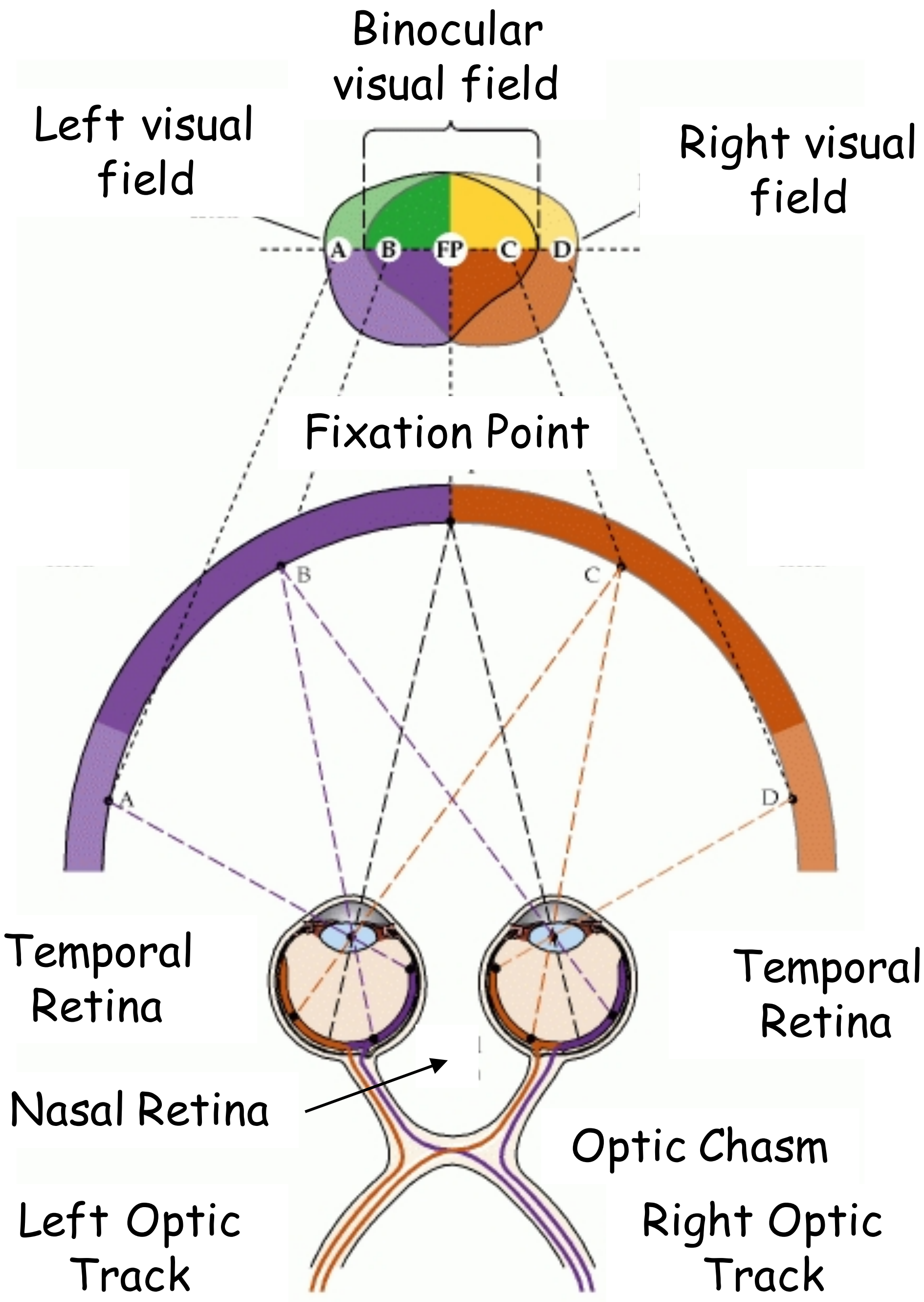
Physiology of the Visual System

We are visually aware of objects in the world by the way in which they absorb, reflect and scatter different wavelengths of light

The eye converts light in the visible spectrum of frequencies (380-750nm) into neural impulses which can then be processed by the brain

The pattern of light which falls on the retina is called the retinal image.

The photoreceptors convert the retinal image into what is known as a neural image which is then sent on to be converted into edges, textures and objects.



Projection

Of binocular field of view onto two retinas and onto optic tracks

National Center for Biotechnology Information: NCBI » Bookshelf » Neuroscience » Sensation and Sensory Processing » Central Visual Pathways » The Retinotopic Representation of the Visual Field

Two Photoreceptors – Rods & Cones

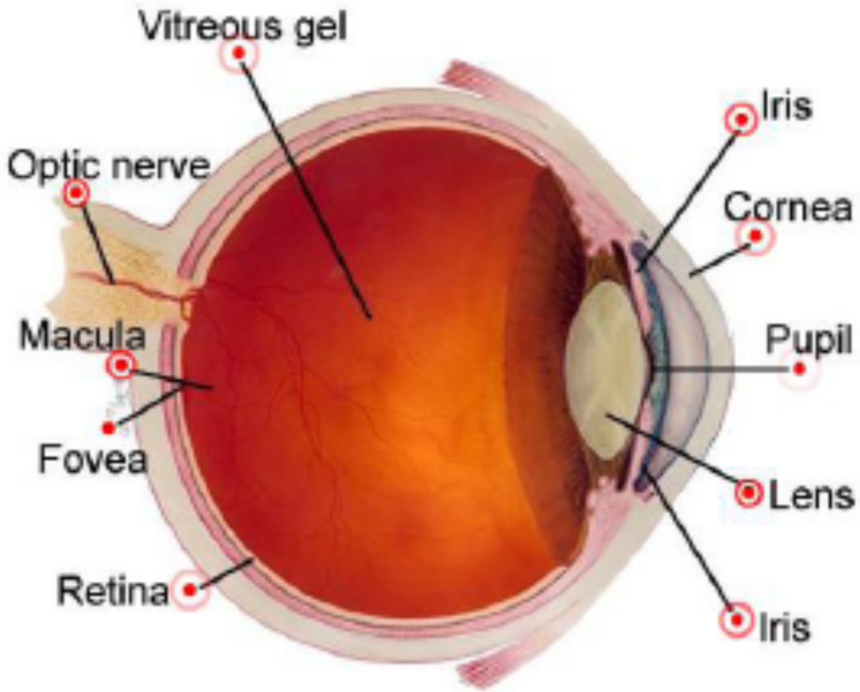
Rods:

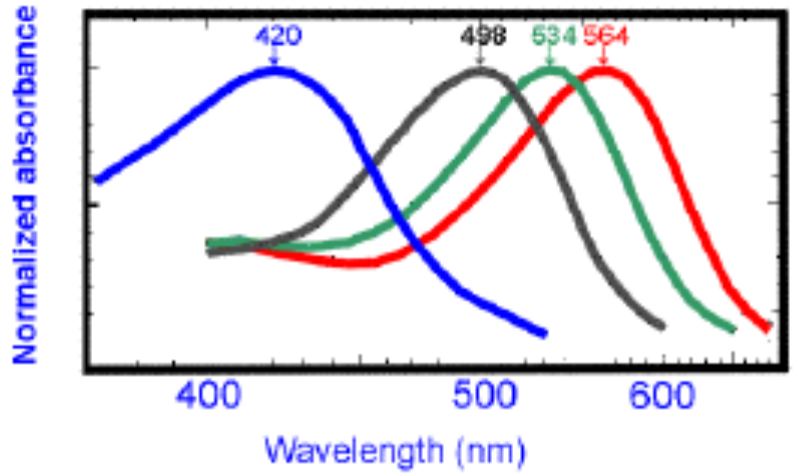
- Most sensitive to lower light levels
- More numerous – 120million Vs 6.5m cones
- Only one "Colour"
- Best at motion detection (great flicker sensitivity) mainly in peripheral vision, no rods in foveal region

Cones:

- Three kinds (usually) VERY roughly most sensitive to one of red, green and blue wavelengths
- Get much higher priority on the neural image (more of the optic nerve)
- Concentrated at the fovea

The Eye and Response



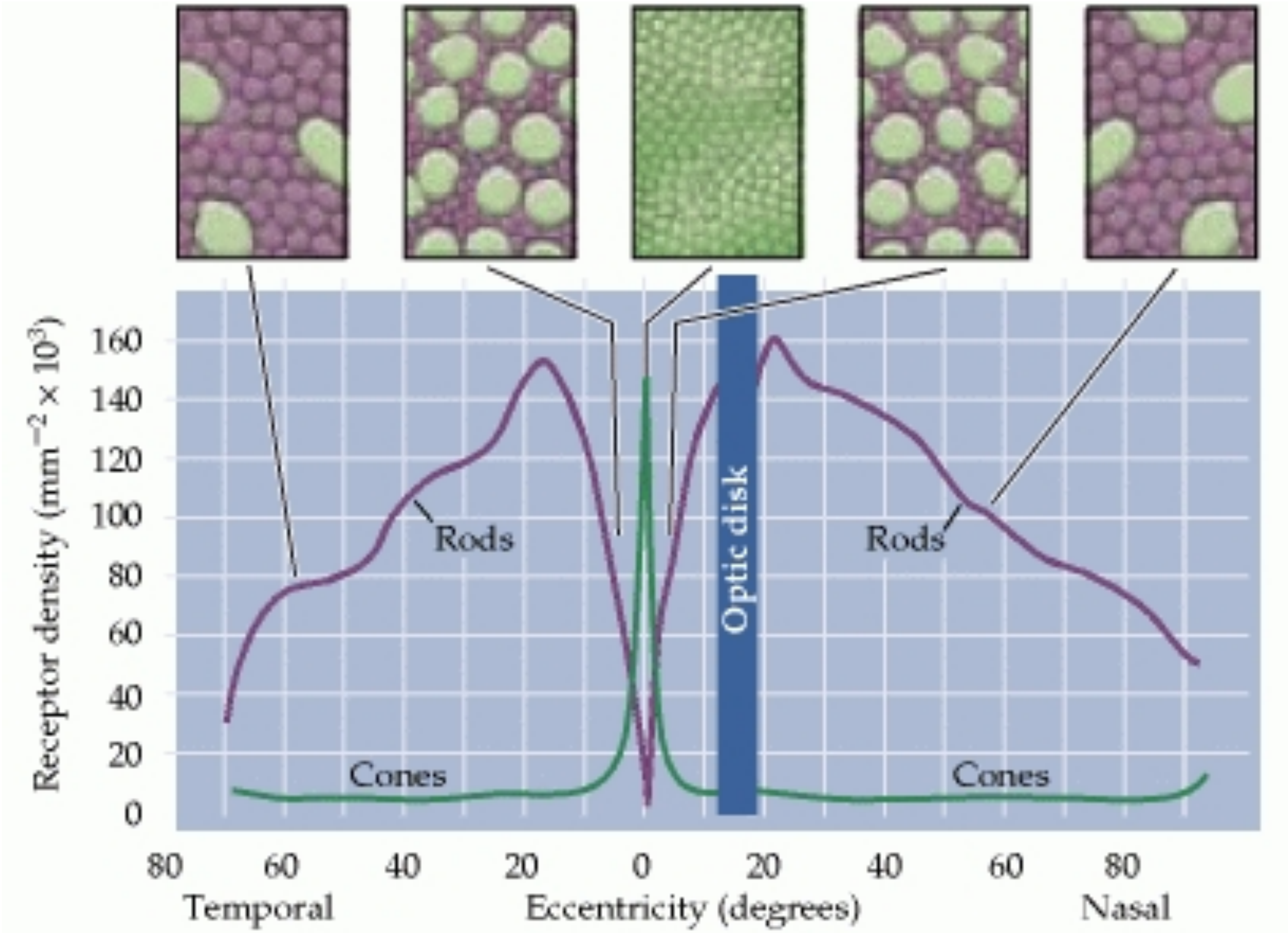


420, 534 and 564nm = Cones
564 = Rods

Eye has Two kinds of photoreceptors: Rods & Cones

Cybernetics

The Human Visual System



Distribution of Rods and Cones In the retina

NCBI » Bookshelf » Neuroscience » Sensation and Sensory Processing » Vision: The Eye »

Visual Performance Characteristics

Smallest resolvable separation in a pattern is approx 0.0083° of visual angle (about 167dpi at 1m)

Field of view:

for one eye = 150° horizontally and 120 ° vertically;

for two eyes = 180° horizontally and 120 ° vertically

Flicker fusion occurs around 60Hz for most people, maximum sensitivity at 10Hz

The perfect visual display will at least exceed these.

Limitations

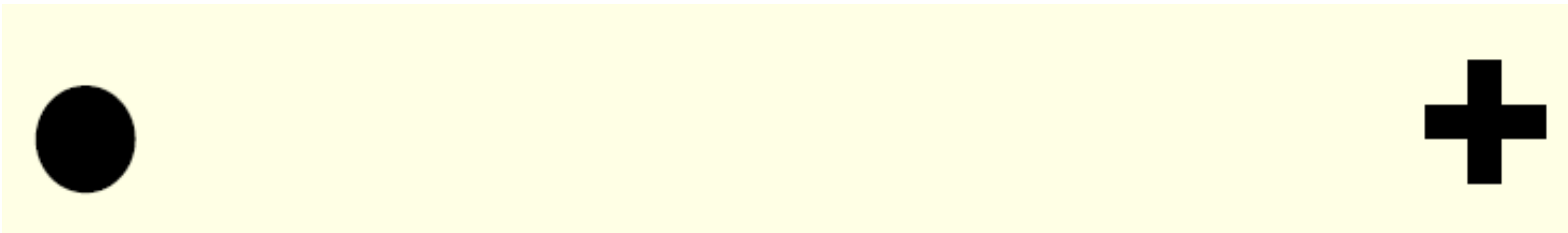
Blind spot –

Cover your right eye. Hold the paper at arm's length.

With your left eye, keep looking at the + while you slowly bring the image closer.

At a certain distance, the dot will disappear – the dot is on the blind spot of your retina.

Try with the left eye



What do you really see?

The one thing we don't see is what's actually there

We interpret the world via very limited sensory organs

Approx a colour by summing narrow bandwidth 'notches'

Miss a huge part of the electromagnetic spectrum

Merge discrete images into continuous motion

Almost no absolute measurement of depth (more later)

Use many assumptions based on expected/learnt phenomena

Optical illusions show where these assumption break down:

http://www.georgemather.com/MotionDemos/BreathingQT.html

But, this turns out to be useful to exploit in VR design

Psychology of the Visual System

Visual perception includes: colour, form, pattern recognition, motion and depth

Depth and motion are more difficult

- there are no (almost) specific receptors for depth so it must be inferred (prone to error)

Depth and motion closely linked

Psychology of the Visual System

Position of an object is found from:

- Relative direction from observer (egocentric direction)
- Distance from observer

Direction is easy, retina is a topographical map of the scene so brain simply maps 2D retinal coordinates to 2D angular coordinates

Depth is harder, no absolute method – have to guess

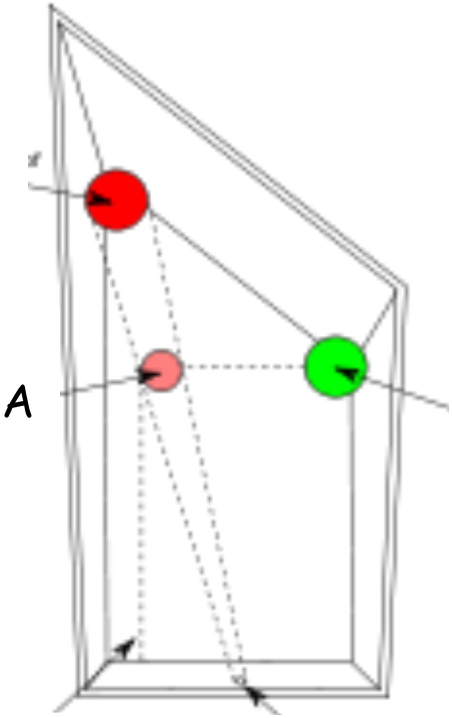
Whole books written about the psychology of vision,

- we'll concentrate on one very small part, very relevant to VR and computer graphics – depth perception

Depth Perception – The Ames Room

How it Works

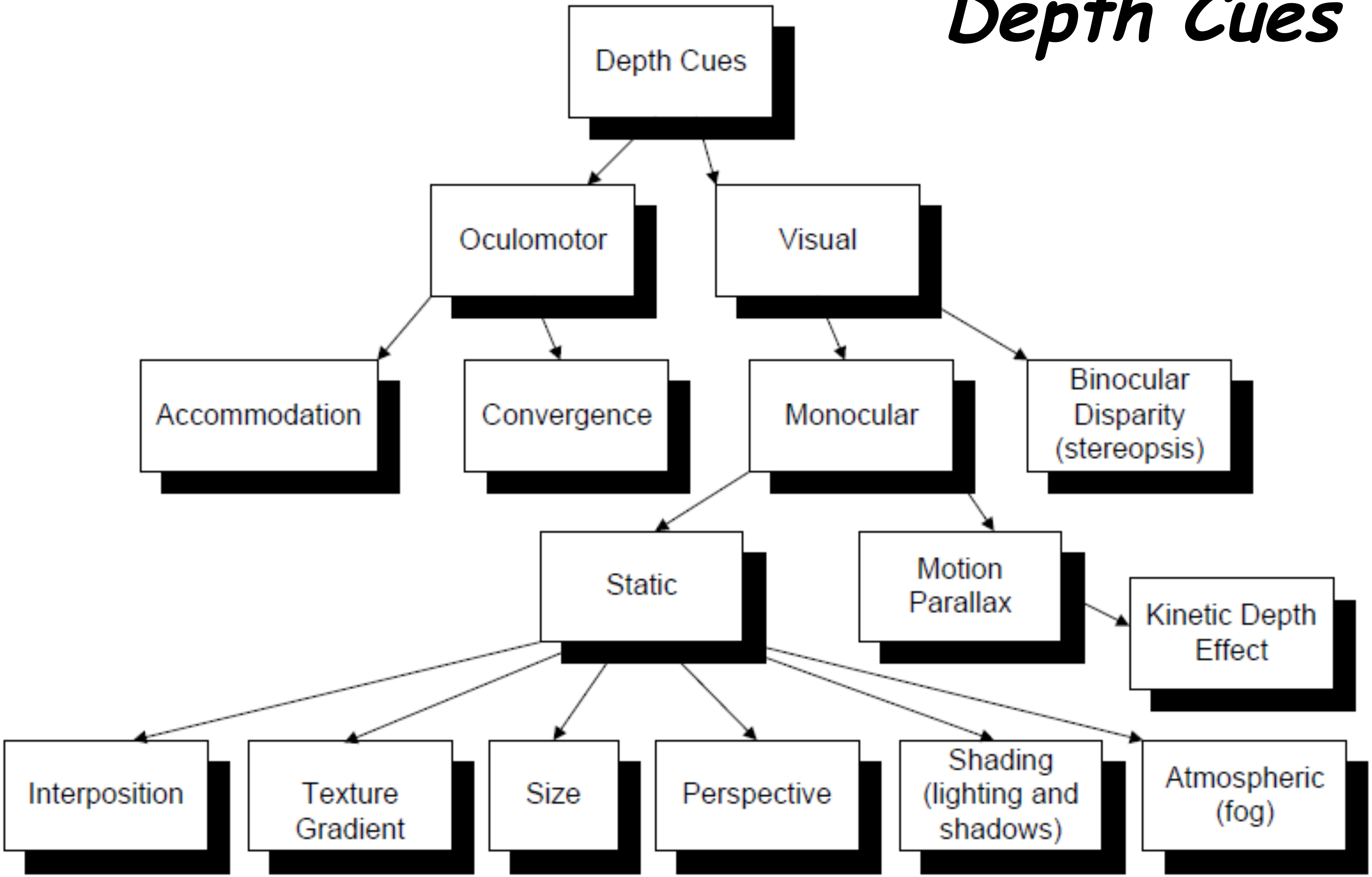
https://en.wikipedia.org/wiki/Ames_room



Used in Lord of the Rings films

Good explanation/demonstration:
http://www.youtube.com/watch?v=Ttd0YjXF0no

Depth Cues



Cybernetics

Depth Cues - Oculomotor

The closest to an absolute measure of depth

Kinaesthetic in origin,

- We move our eyes, using muscles
- Information from muscles to brain give depth cues

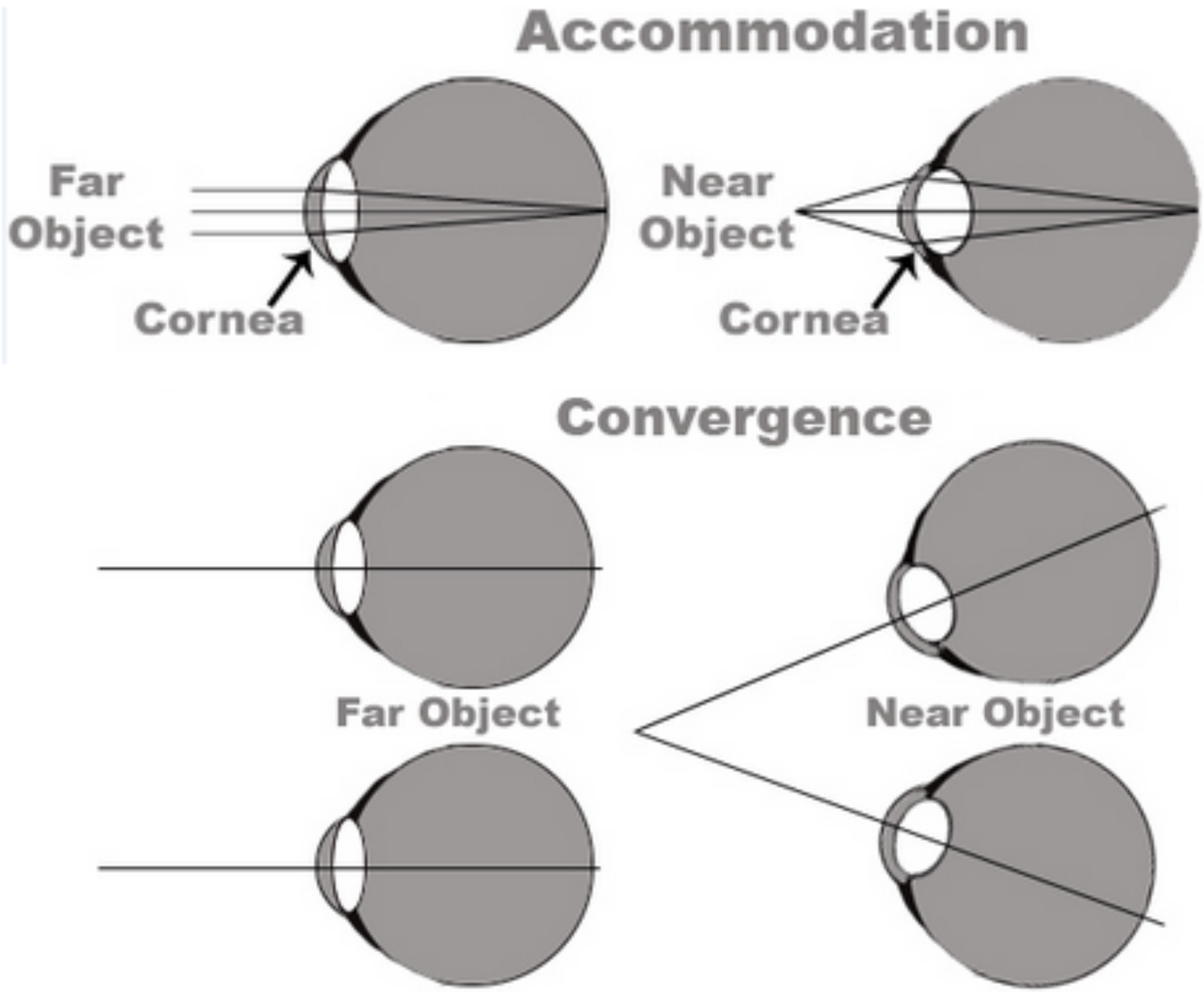
Accommodation: how much the ciliary muscles are pulling on the crystalline lens to focus the image

- (almost useless as a depth cue)

Convergence: the orientation of the eyes,

- i.e. the closer an object the more the eyes rotate towards each other.

Figures

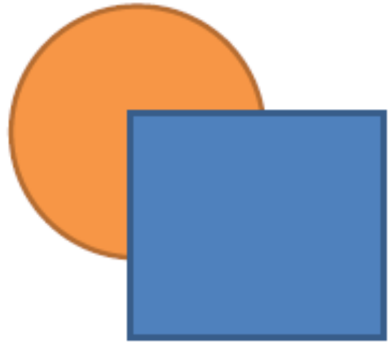


http://ucalgary.ca/pip369/mod4/depthperception/oculomotor

Depth Cues – Static, Monocular

Interposition:

- When one object partially obscures another
- Simple but strong, will override most others



Size:

- Expected depth can also affect perceived size

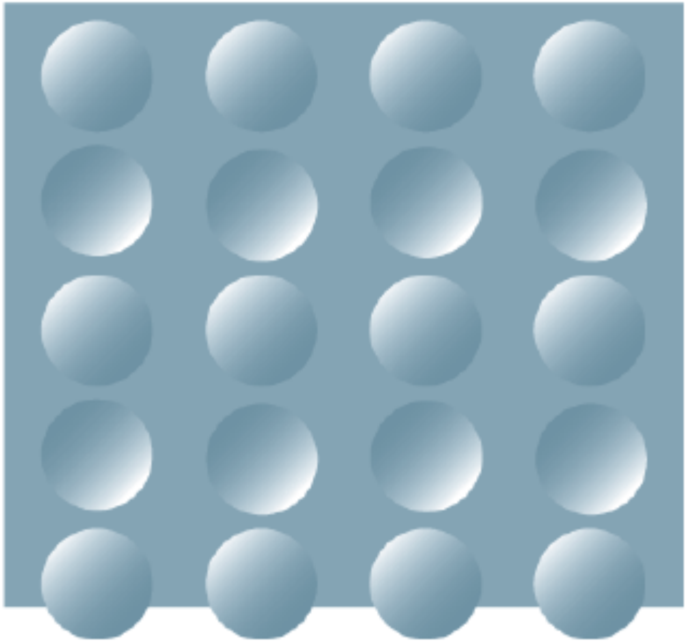
Linear perspective/Texture Gradient

- Both closely linked
- Parallel lines converge as they get farther away
- Horizontal lines get closer

Depth Cues – Static, Monocular

Shading:

Includes Lighting and Shadowing

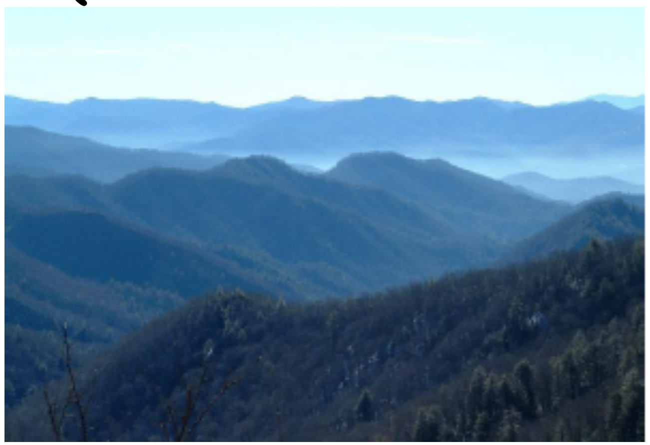


Atmospheric

Distant objects become less clear

- air pollution, fog etc.

(also bluer – used by artists)



Shadows - give relative position between objects

Depth Cues – Stereopsis (seeing solid)

Retinal Disparity

- Each eye gets a slightly different view
- Brain fuses images together to create 'solid' objects
- Less effective with distance

http://cmp.felk.cvut.cz/demos/Stereo/New/Matching/cedar-right.png

Depth Cues – Dynamic

Motion Parallax

- When objects or the viewer are in motion, closer objects move faster across the retina than those farther away
- When a single eye moves approx the distance between the eyes (60-70mm) the depth cue is equivalent to binocular disparity
- This effect is quickly lost when scene stops moving

Kinetic Depth

Kinetic Depth effect (Similar to parallax)

- Objects can seem 3D when moving and 2D when stationary.
- Think of a glass sphere with dots painted on it.
- http://www.youtube.com/watch?v=XGfMB76U5ts
- http://www.georgemather.com/MotionDemos/KDEQT.html

These are examples of the human mind attempting to make sense of the world from limited and incomplete sensory input

Motion Perception

Motion is a very Primary quality* – it can be perceived visually, audibly, tactilely, through smell and even balance

In the real world objects move smoothly and continuously.

In the virtual world (and cinema, TV etc) they 'jump'

For all the senses a threshold exists where two successive stimuli are perceived as one continuous stimulus

- For vision this is known as flicker fusion ~17ms*
- For audition this is know as flutter fusion ~10ms*
- For touch this is known as vibration and is around 5ms*

* Fraisse, P., Time And Rhythm Perception, in Handbook Of Perception. 1978, Academic Press. p. 203-247.

Motion Perception – Apparent Motion

BUT fusion is not necessary for motion to occur,
  it is preferable though

Apparent Motion occurs when the time delay between successive stimuli is lower than the fusion threshold but the brain still perceives motion

Lots of varieties: Short range, long range, transformational, biological

Some of theses effects work for touch and sound too.

Examples here:
  http://www.georgemather.com/MotionMP4.html

Summary Questions

1. A VR interface is to be designed to train jet fighter pilots. There is some debate as to whether the interface should have stereoscopic 3D vision. What are some of the arguments for and against in terms of depth perception and the pilots actions?
2. Many TVs are now being sold as '3D Ready'. Why might it be said that all TVs are 3D?

IMPORTANT: These questions are NOT compulsory and are NOT part of the course mark. They are for you to test your understanding of the lecture material.

Visual Display Technologies

Most displays can be categorised as:

- Desktop
- Wearable
- Projection

Most of the 3D cues from last lecture are part of the 2D image, stereoscopy needs dedicated hardware

- Generating most good 3D cues therefore a software/rendering problem
- Why We might refer to any TV or monitor as “3D ready”

Common stereo techniques discussed next, others exist

Stereoscopic Display Techniques

A different image must be presented to each eye

If both eyes viewing same display surface need to selectively block or transmit light

Most single surface displays have some ghosting (crosstalk)

- when one eye sees parts of image intended for other
- Can cause discomfort, loss of presence / stereo fusion
- In practice single display systems are rarely completely free from ghosting

Three main varieties: active, passive and autosteroscopic

Active Stereo

Left/Right image are alternated on display surface. LCD 'shutter' glasses block each eye alternately

CrystalEyes 3, StereoGraphics

3D Vision
Nvidia

Ghosting affected by:

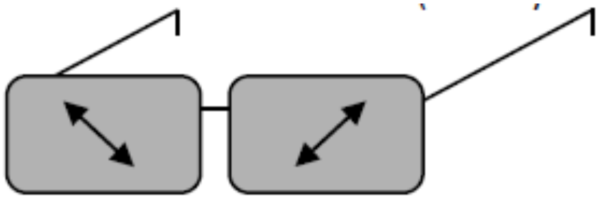
- Display fade time
- Light blocked by LCD
- Synchronisation of glasses and display

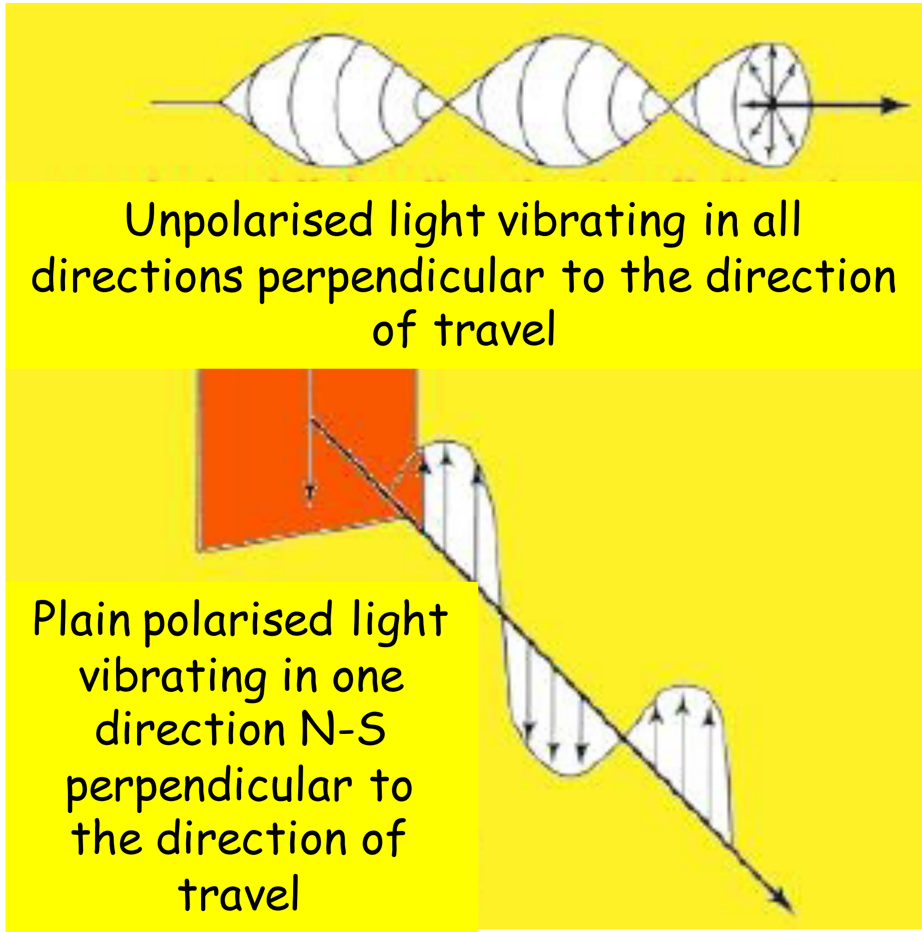
Passive Stereo - Polarisation

Many passive stereo techniques
  Polarisation and Frequency
  Multiplexing most common

Viewer wears glasses : each
  lens allows only one direction
  of polarised light to pass

Display transmits each L/R
  image with diff orientation
  of light polarisation





http://www.mineralatlas.com/Optical%20crystallography/properties%20of%20light.htm

On Polarisation

Linear polarisation:

- Cheaper/easier to manufacture with high % light absorption/transmission
- User can't rotate head

Circular Polarisation

- Often more ghosting
- Invariant of viewer orientation
- Display surface must preserve polarisation (usually requires special coating)

Passive Stereo–Frequency Multiplexing

Frequency Multiplexing –

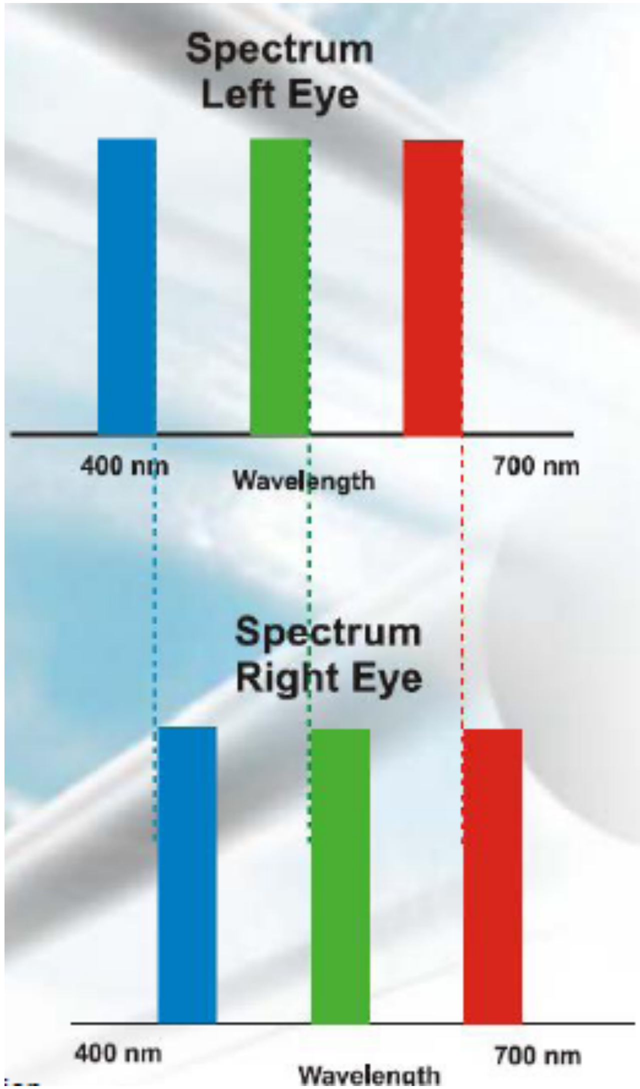
Old style Anaglyph, newer Infitec.

Block diff parts of spectrum

Invariant of head rotation.

- Anaglyph loses colour info.
- Infitec can preserve colour if intensities correctly adjusted
- Any display surface can be used

http://kids.usachoice.net/Tiger-Anaglyph.JPG



Autostereoscopic Displays

Don't use glasses
lenses/barriers make sure each eye sees a different part of the display

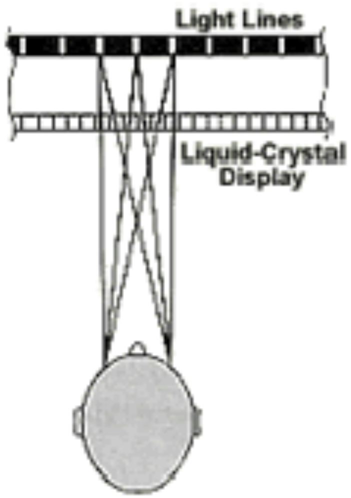


Figure 2

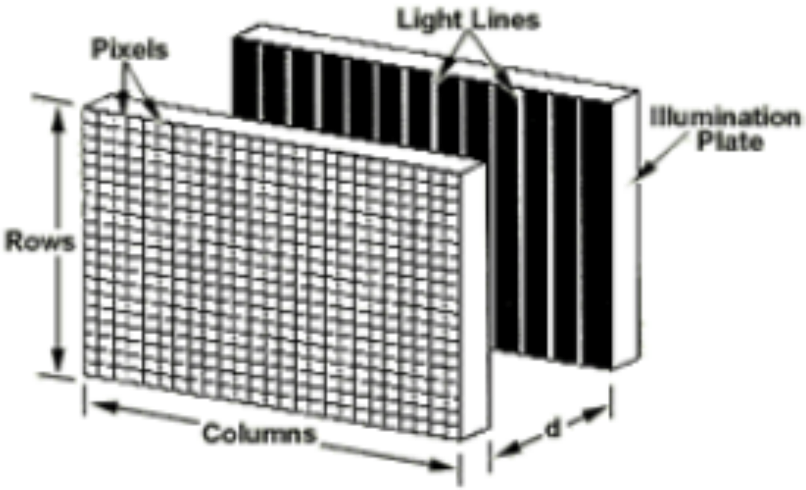
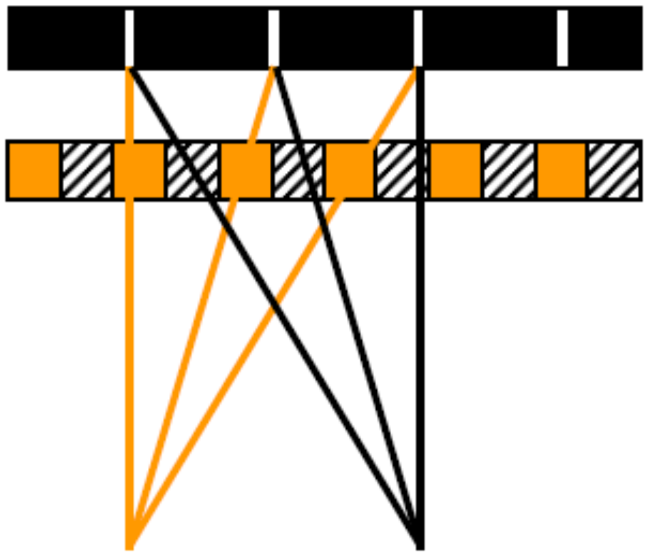


Figure 1

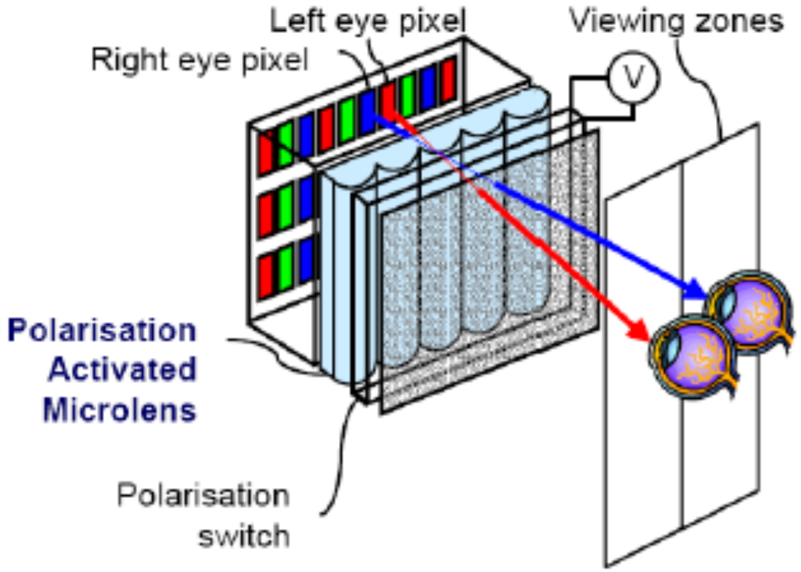


http://dti3d.com/About/dti_technology.htm

Autostereoscopic – Fixed Barriers

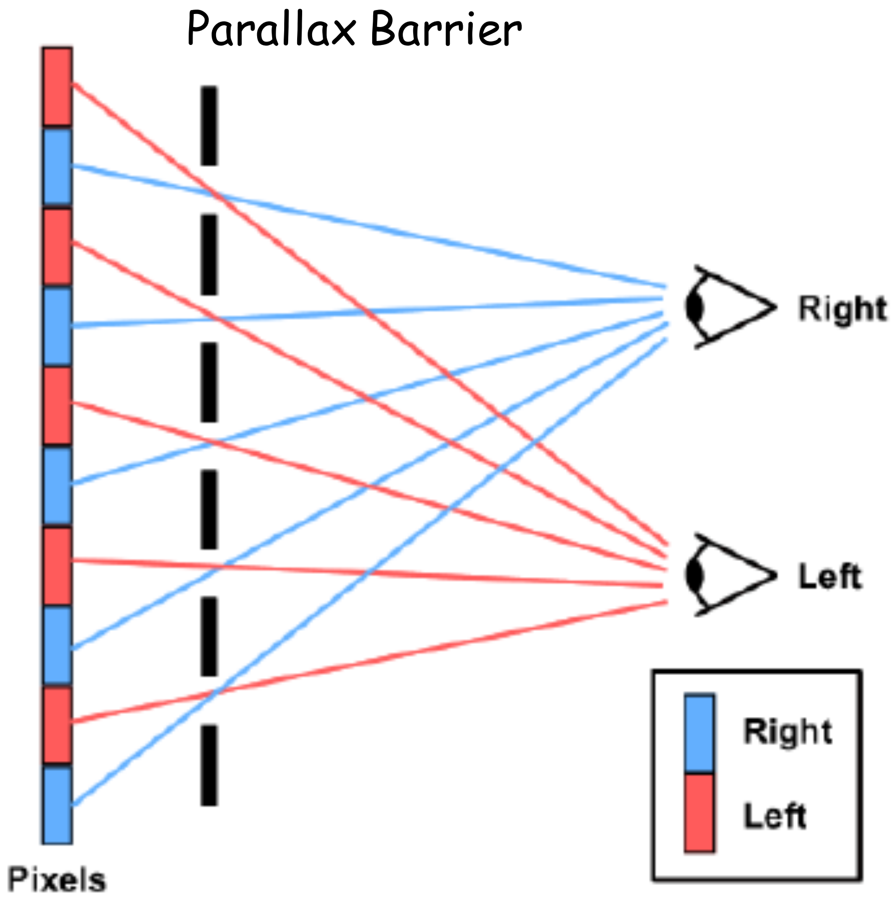
Have a limited viewing angle and 'sweet spots'

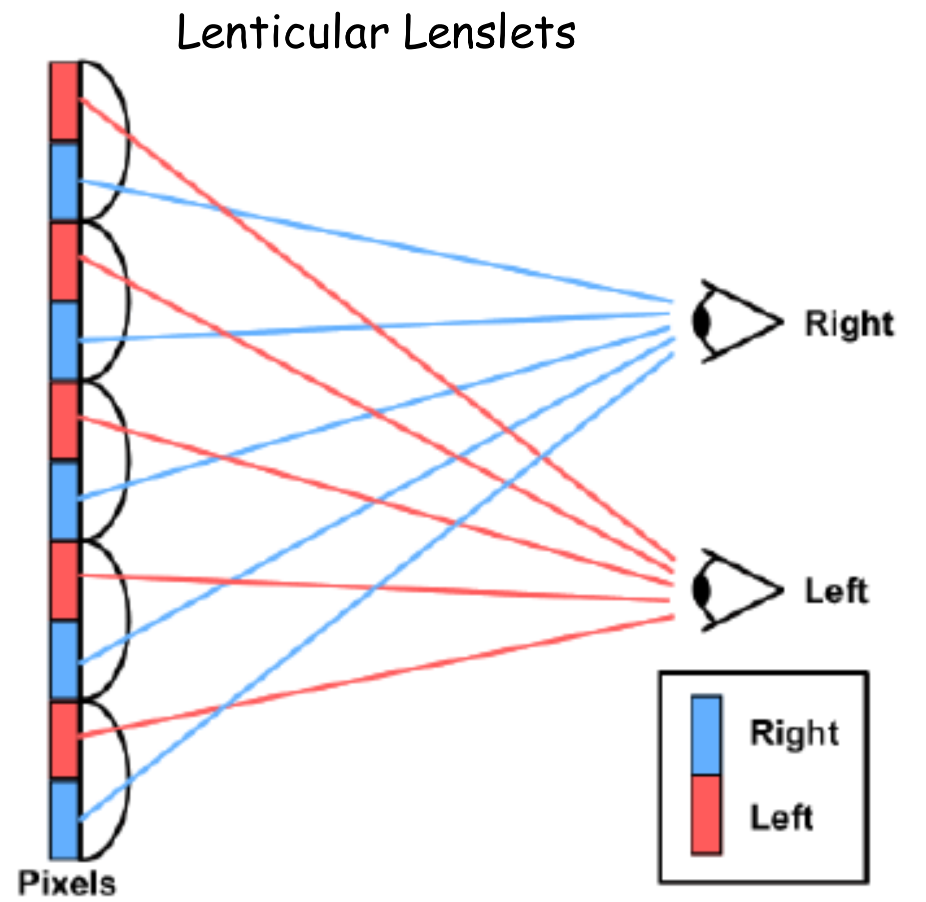
- Can sometimes allow multiple viewing positions (look around) by trading off resolution
- Don't scale well, better for small displays



From www.vrtechnology.org

AS – fixed lenses/barriers



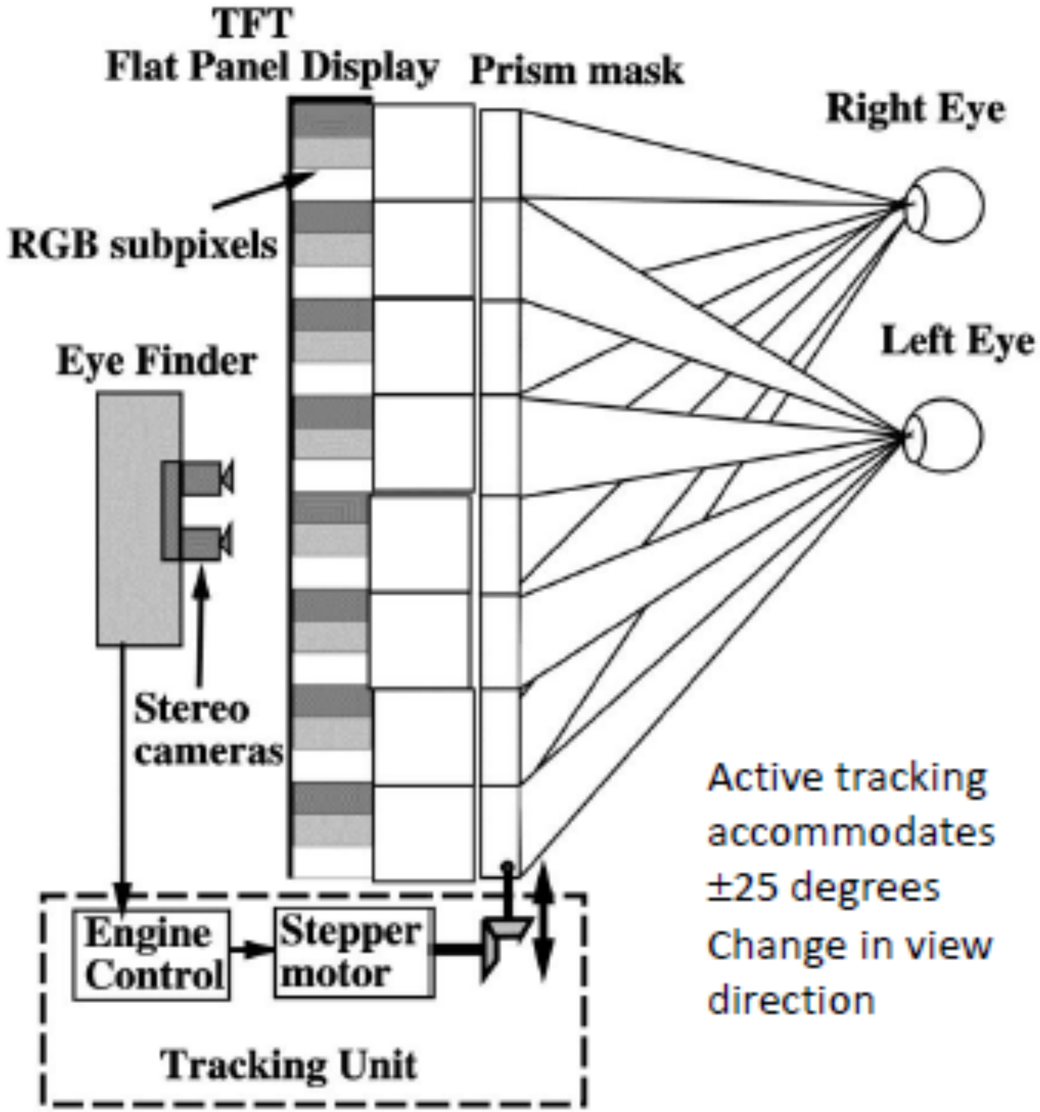


From www.3dforums.com

AS - Movable Barriers

To increase viewing angle : track user's eye position and adjust barrier to match

- Get 'look around' effect
- Only one viewer at a time



Active tracking accommodates ±25 degrees

Change in view direction

Cybernetics

Comparing Visual Displays

Lots of criteria for selecting/comparing performance:

Technological:

- Field of View/Field of Regard
- Resolution
- Update Rate
- Interface with tracking methods
- Associability with other sense displays - i.e. integration with audio and haptic displays

continued

Continued

Ergonomic:

- User mobility
- Environment requirements
- Throughput - the number of people who can experience the VE in a given time
- Safety
- time.
- Encumbrance

Lots of others:

- Portability
- Cost

Desktop displays

Cheap, common, accessible
- Small FoV/FoR
- Use all the stereo techniques discussed

http://www.panoramtech.com

Linear Polarised:

http://www.planar3d.com/

Shutter Glasses:

http://www.nvidia.com

Autostereoscopic:

http://www.seereal.com/

Circular Polarised (Row Interlaced):

http://www.zalman.co.kr

Wearable Displays

Just considering head mounted. Also have boom mounted and hand held.

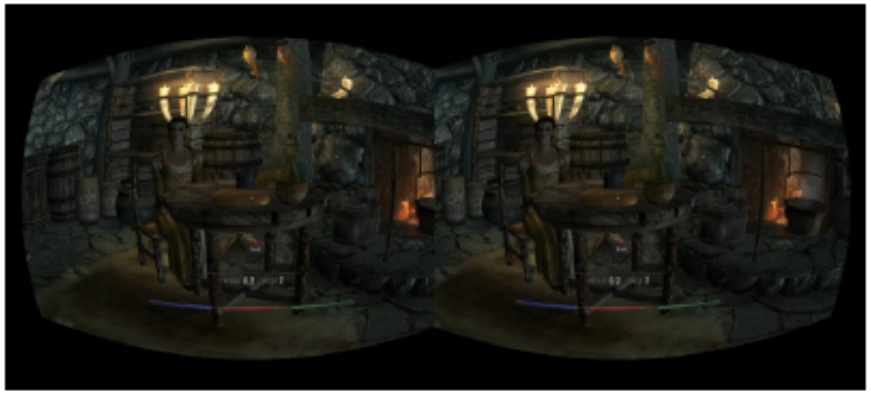
- More expensive than desktop, especially for high quality (£10k++)
  - OculusRift may change this!
- Separate display for each eye
- Optics magnify image and change focal length (1-5m)
- Resolution generally low
- Usually limited field of view <60°
- Trade off FoV <-> Resolution
- Excellent (complete?) Field of Regard

Visette 45, Cybermind

Wearable Displays

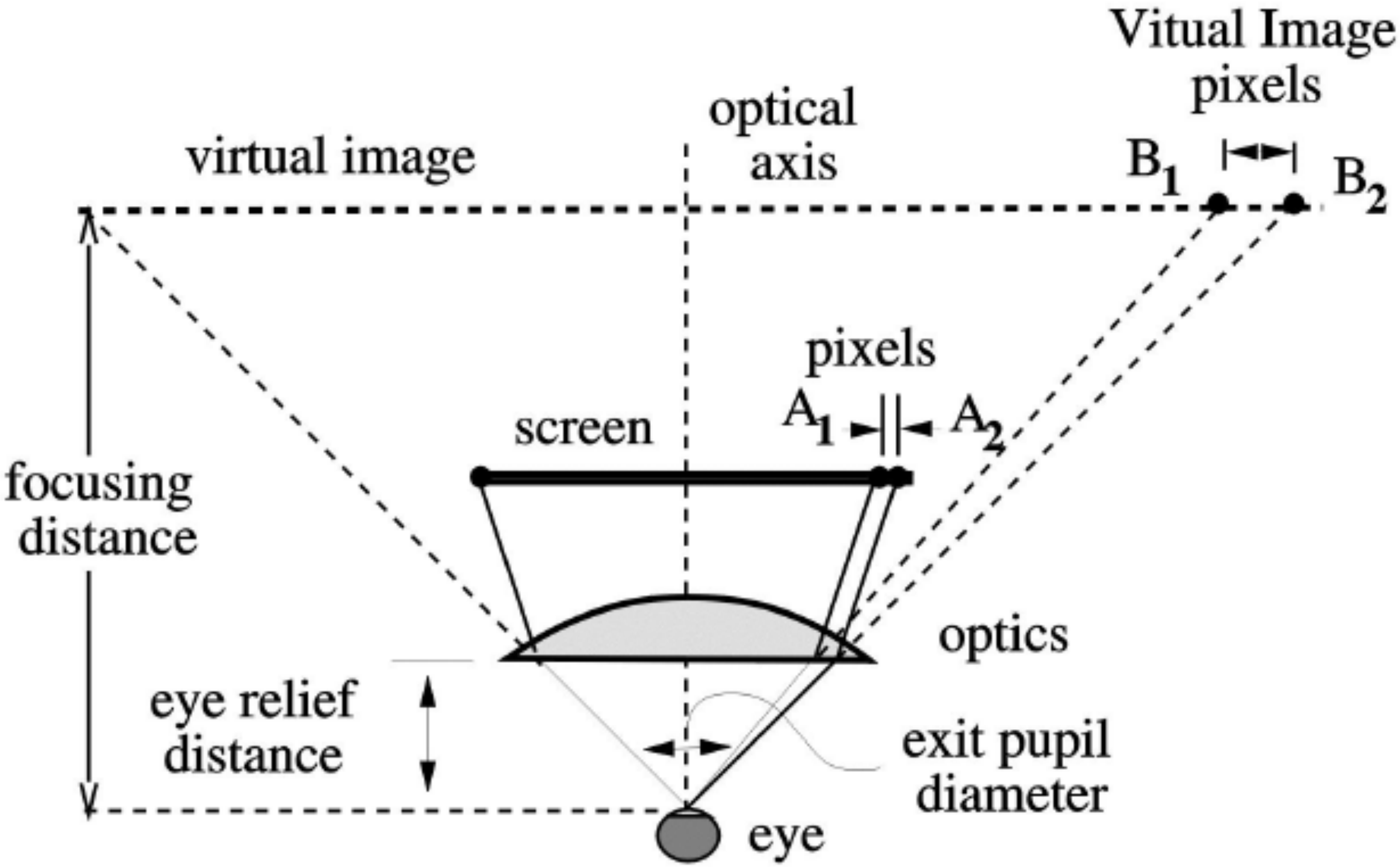
Real world normally blocked from view

- Weight & comfort a big issue
- Low latency tracking essential.
- Can't see own body
- Balance and movement affected
- Unnatural body positions often adopted
- Distance travelled often underestimated by up to 25%, unclear why



Head Mounted Display

A simplified model of the optics (from www.vrtechnology.org)

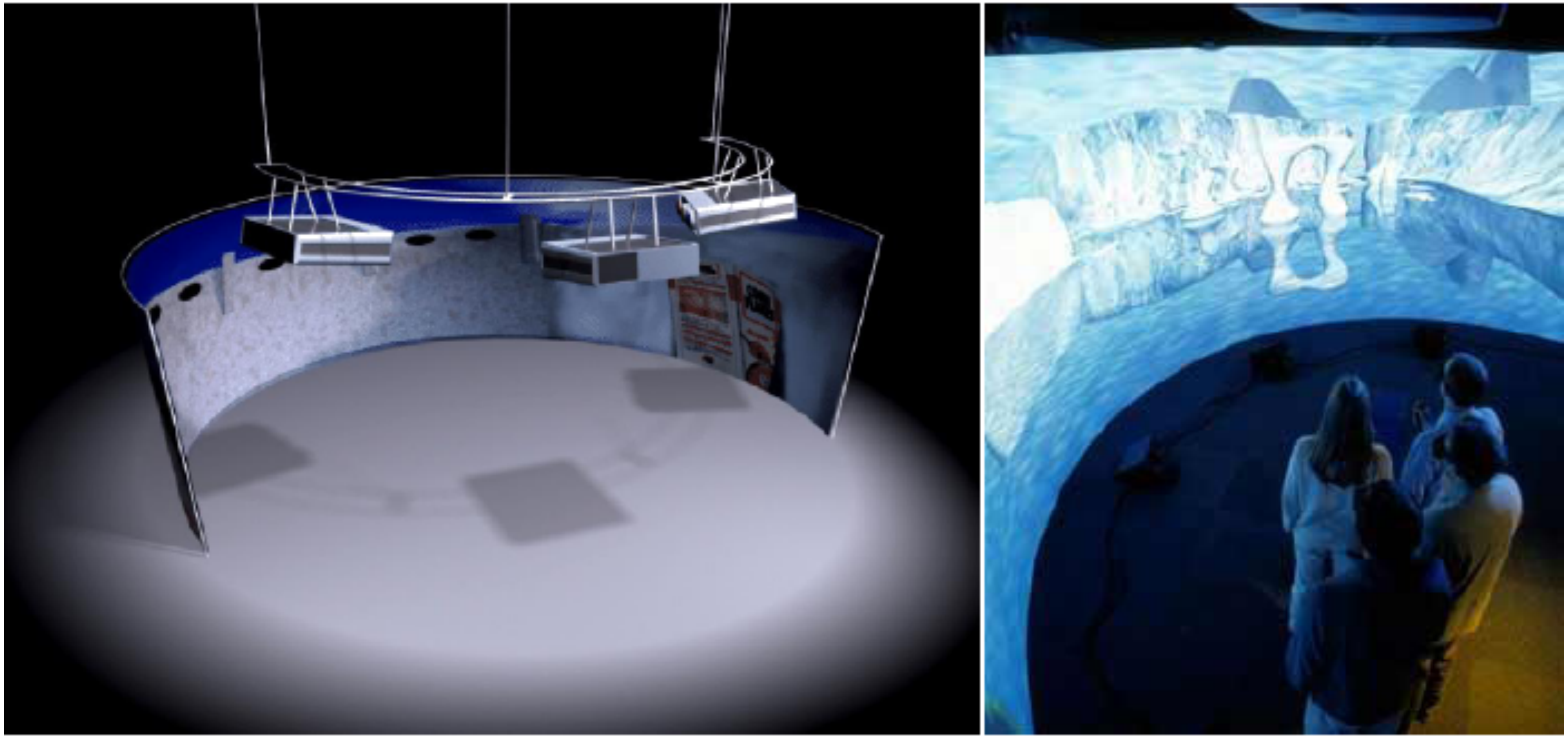


Projection Displays

One or more projectors or screens

- Large area – good field of View
- 2D or Stereo (Linear, Circular or Infitec most common)
- Usually rear projected to prevent users' shadows
- Can see own body – good for presence and natural movement/interaction
- Multiple users can share environment (configurations can affect this i.e. head tracking)
- Different configurations in use:
    - Panoramic, Vision Domes, Spheres, Power Walls

Panoramic Displays



i-Cone, Fraunhofer IMK

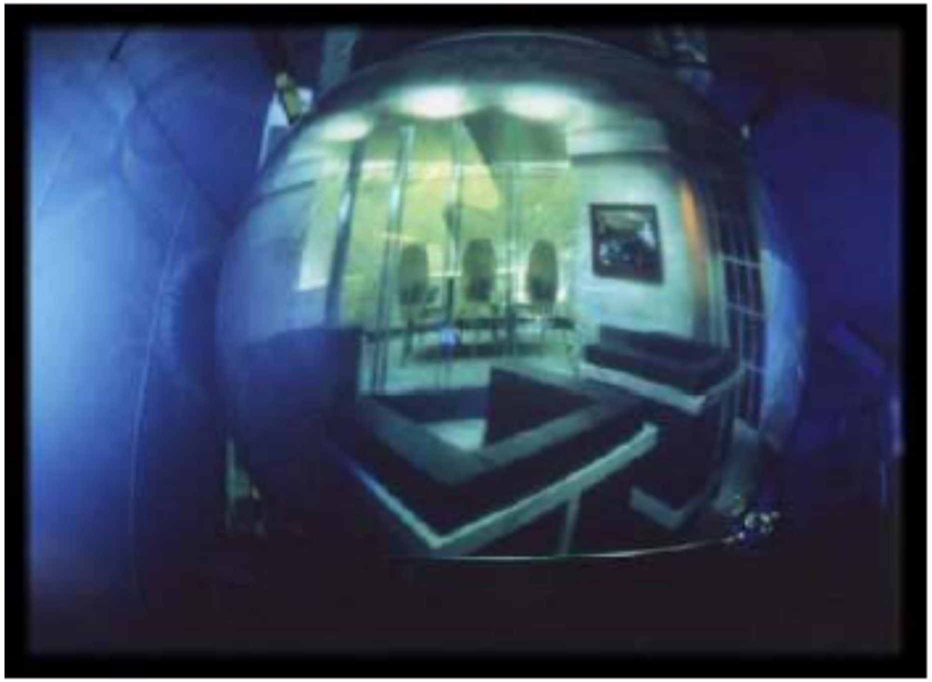
Cybernetics

Vision Domes

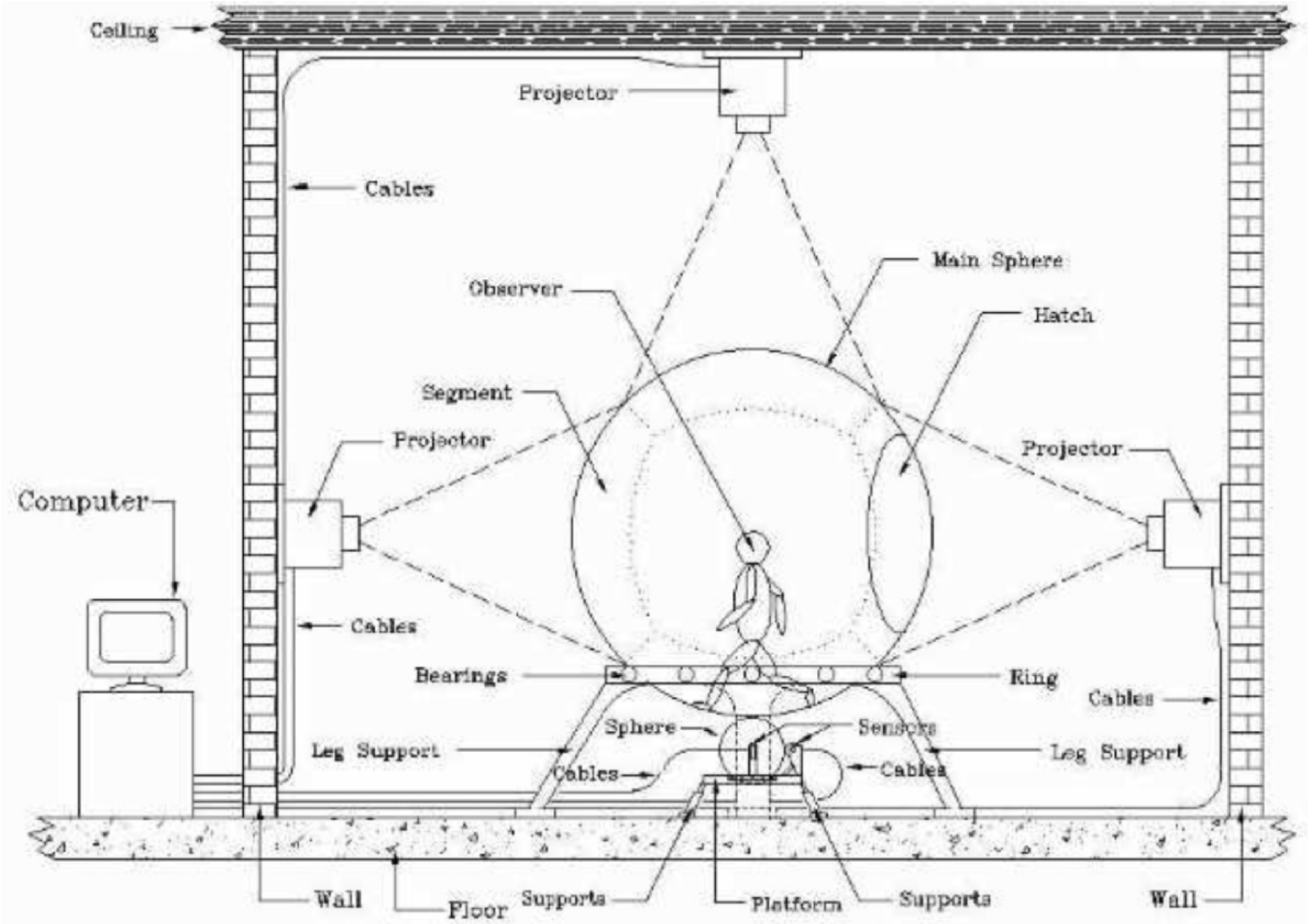
http://www.avrrc.lboro.ac.uk

http://www.cg.tuwien.ac.at

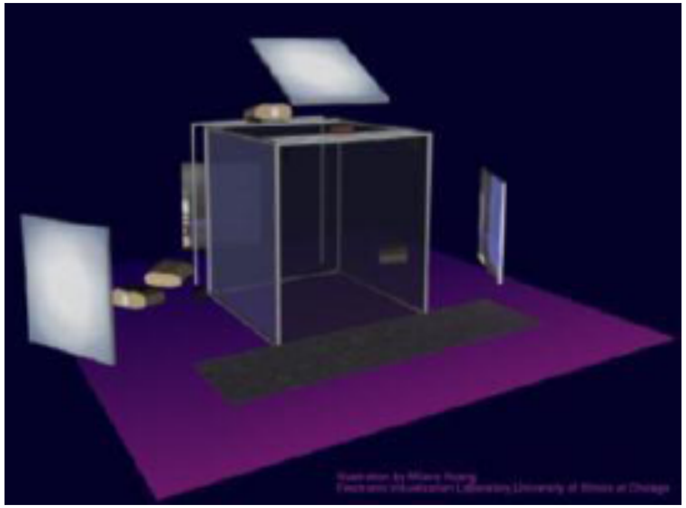
Spheres

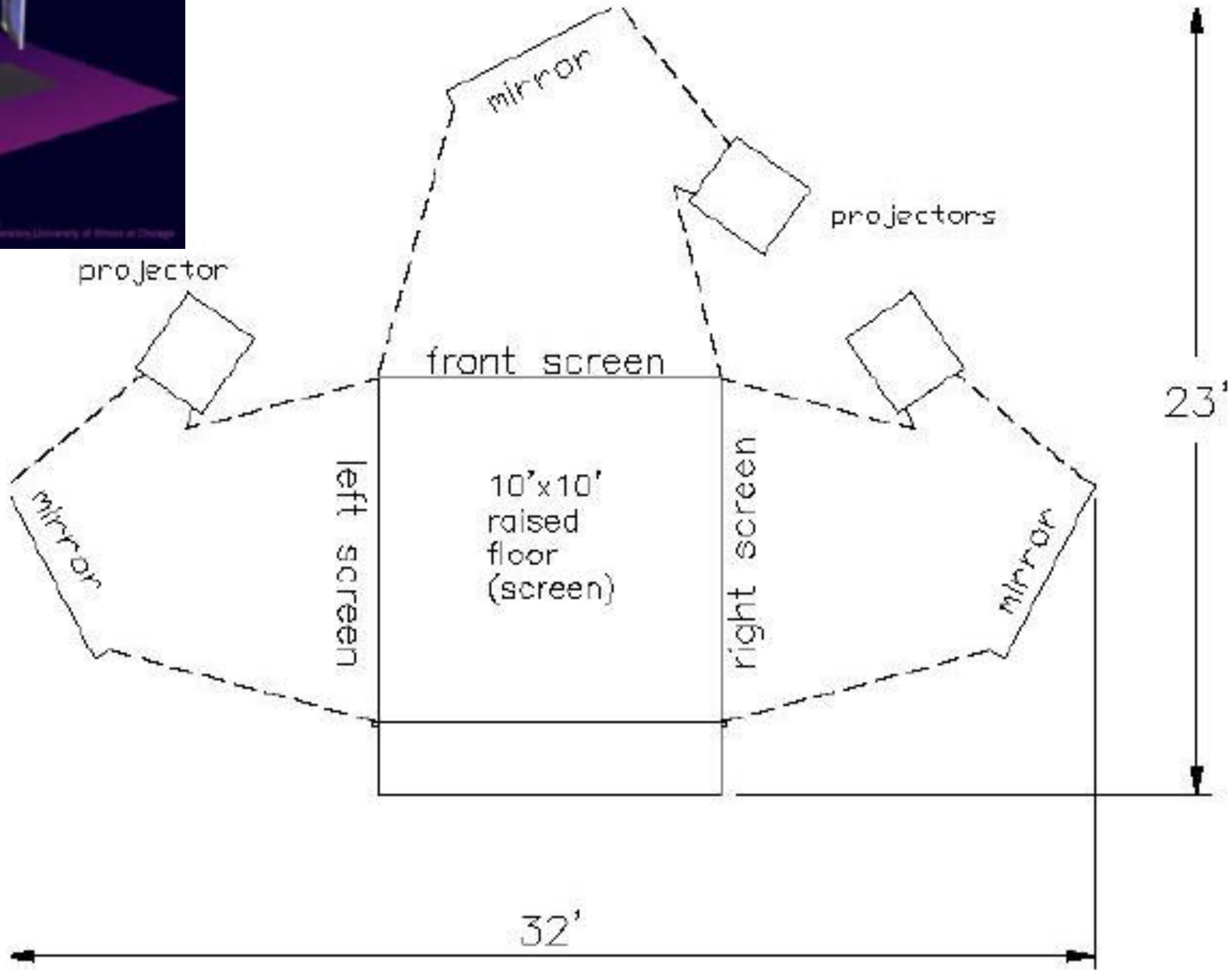


Cybersphere
http://www.vr-systems.ndtilda.co.uk/sphere1.htm


Ceiling
Projector
Cables
Main Sphere
Observer
Hatch
Segment
Projector
Projector
Computer
Cables
Bearings
Ring
Cables
Sphere
Sensors
Leg Support
Leg Support
Cables
Cables
Wall
Floor
Supports
Platform
Supports
Wall

Cubic Displays





Computer Assisted Virtual Environment

Or CAVE Assisted Virtual Environment

Has 3 to 6 large screens

Puts user in a room for visual immersion

Usually driven by a one or more powerful graphics engines

Can put equipment into room to aid realism

  Eg shopping trolley in model of supermarket

Can have multiple people

Advantages of CAVE

Provides high resolution and large FOV

Uses peripheral vision

Use pair of light weight shutter glasses for stereo viewing

User has freedom to move about the device

Has space to place props (cockpit etc.)

Environment is not evasive

Real and virtual objects can be mixed

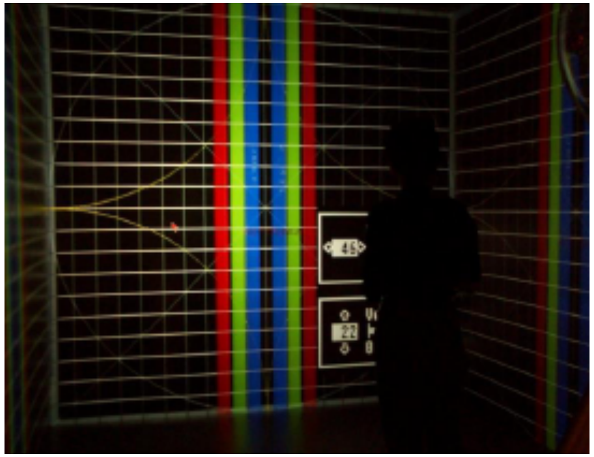
A group of people can inhabit the space

Disadvantages

Very expensive (approximately 1 million dollars)
Requires a large amount of physical space
Projector calibration must be maintained
Only 1-2 users can be head tracked
Stereo viewing can be problematic
Physical objects can get in the way of graphical objects

Projection Displays - Issues

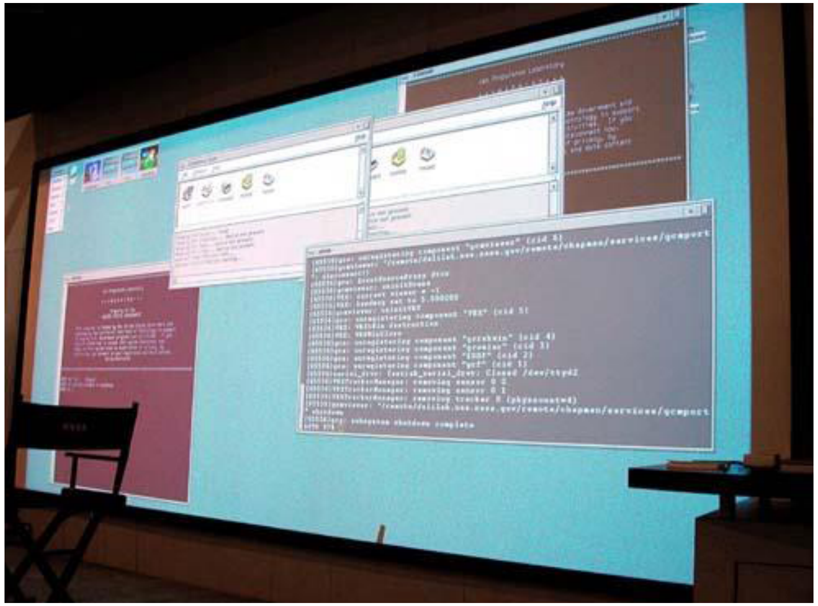
Multiple projectors must have:
- Geometric continuity
  - (correct alignment)
- Photometric continuity
  - (consistent brightness and colour)

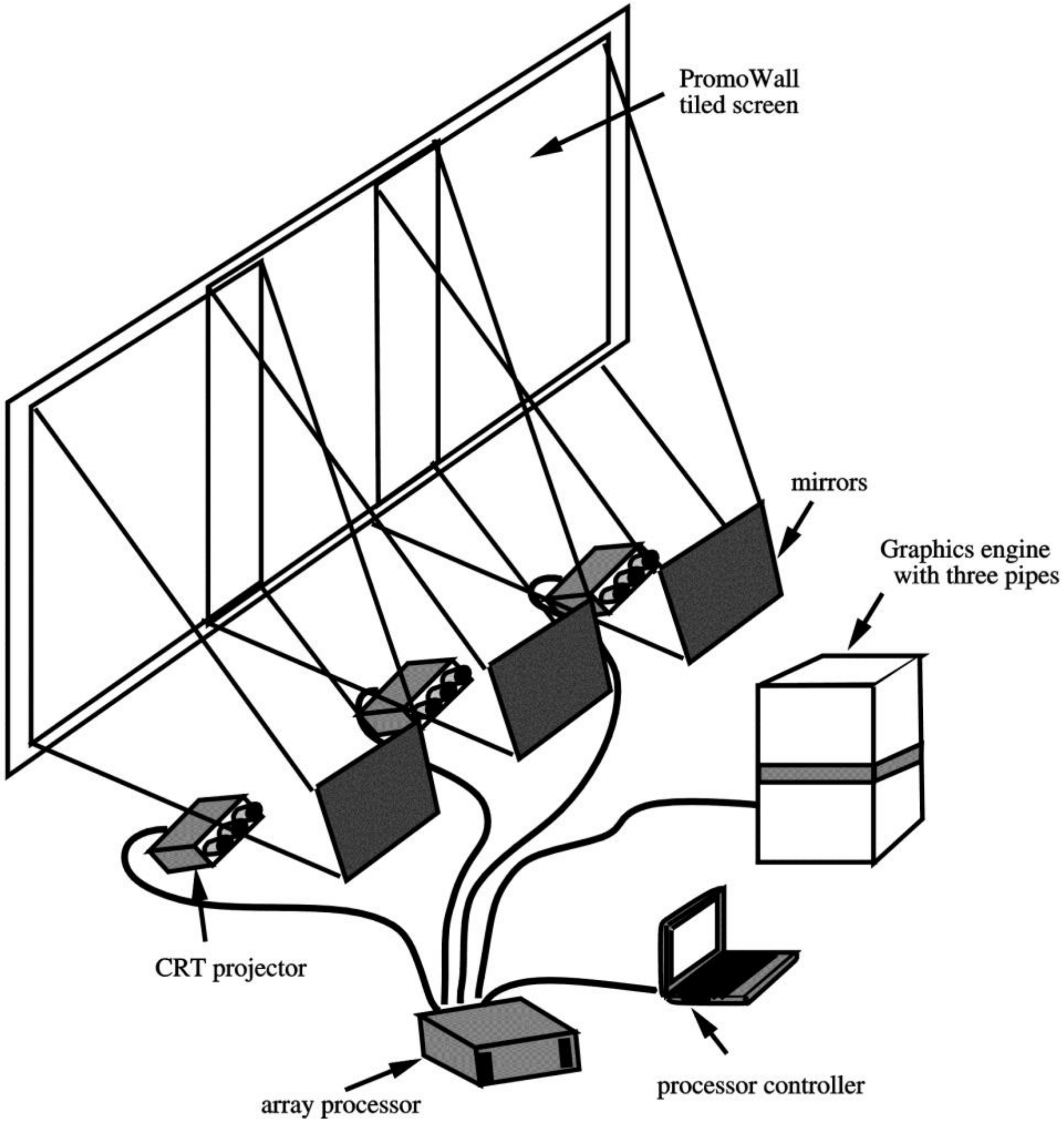


Wall Displays

To keep resolution high, multiple projectors usually tiled

- Continuity particularly important
- Where images overlap, need to adjust brightness
  - Can adjust brightness in software or hardware (blend plates)





From www.vrtechnology.org

Wall Displays – Tiling and Blending

Summary

Providing a good visual representation of a virtual world is important

There are numerous technological solutions

It is an active area of work

As is shown by the 2016 Consumer Electronics Show, where VR features – see

http://www.bbc.co.uk/news/technology-35192737

Next week we start to consider hearing virtual worlds

Auditory Perception

We have explored the ways in which virtual worlds can be displayed

Now we consider how they can be heard

Again we look at how we perceive the real world

this time hearing it

And then look at the technologies to allow us to hear virtually.

Transmission of sound

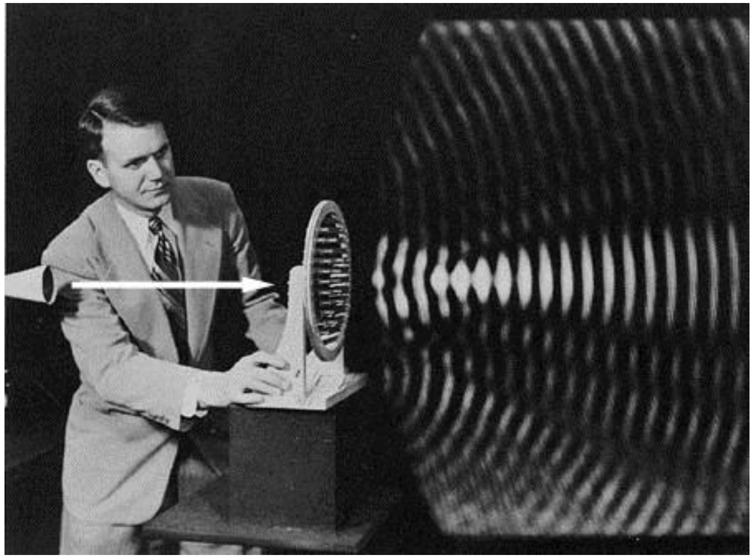
Sound travels as a pressure wave through any medium dense enough to sustain it.

Speed related to medium's density (340m/s in air at sea level)

Intensity drops off with distance (Inverse Square Law)

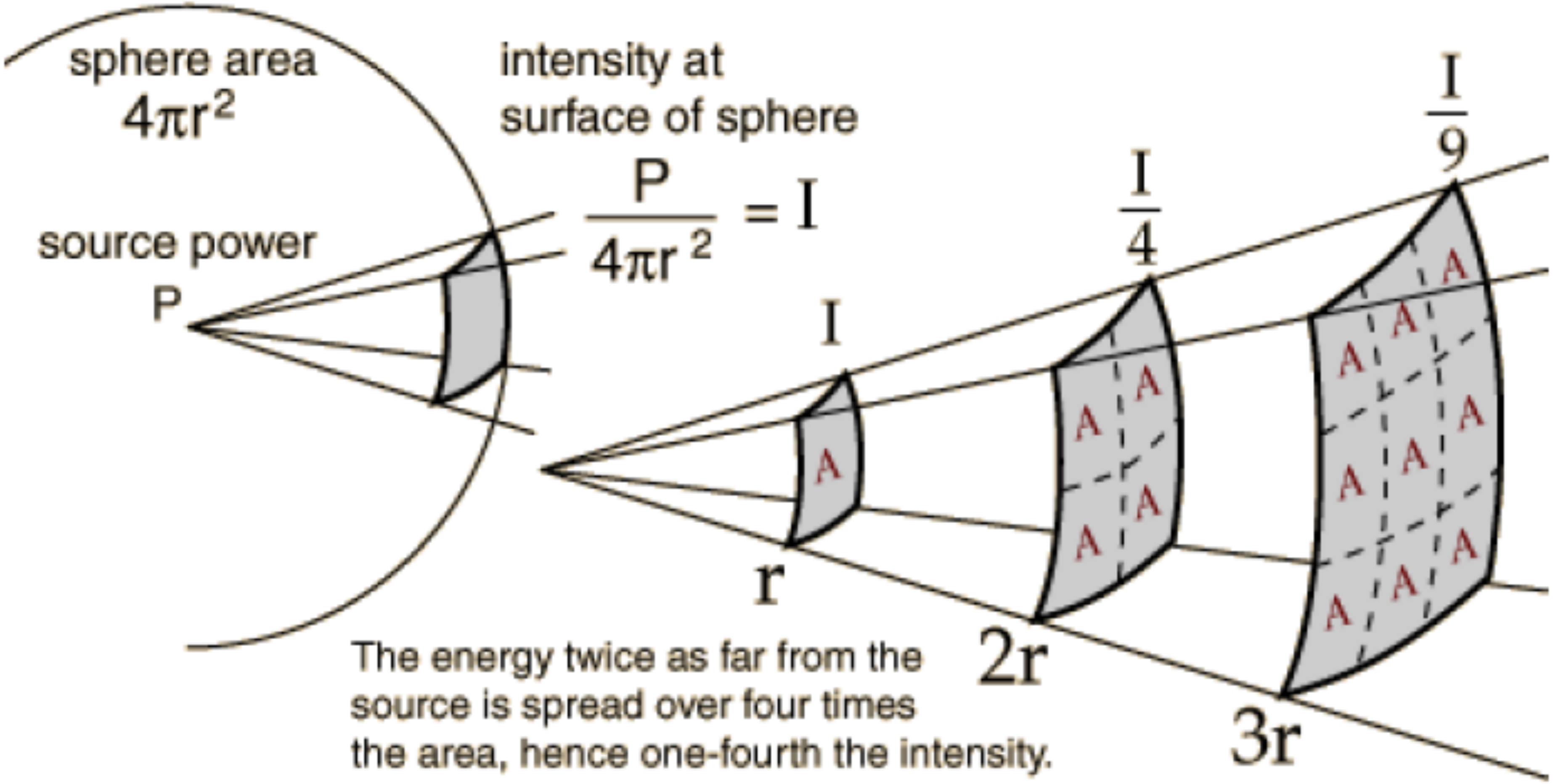
Intensity loss related frequency^2

High Freq attenuate more



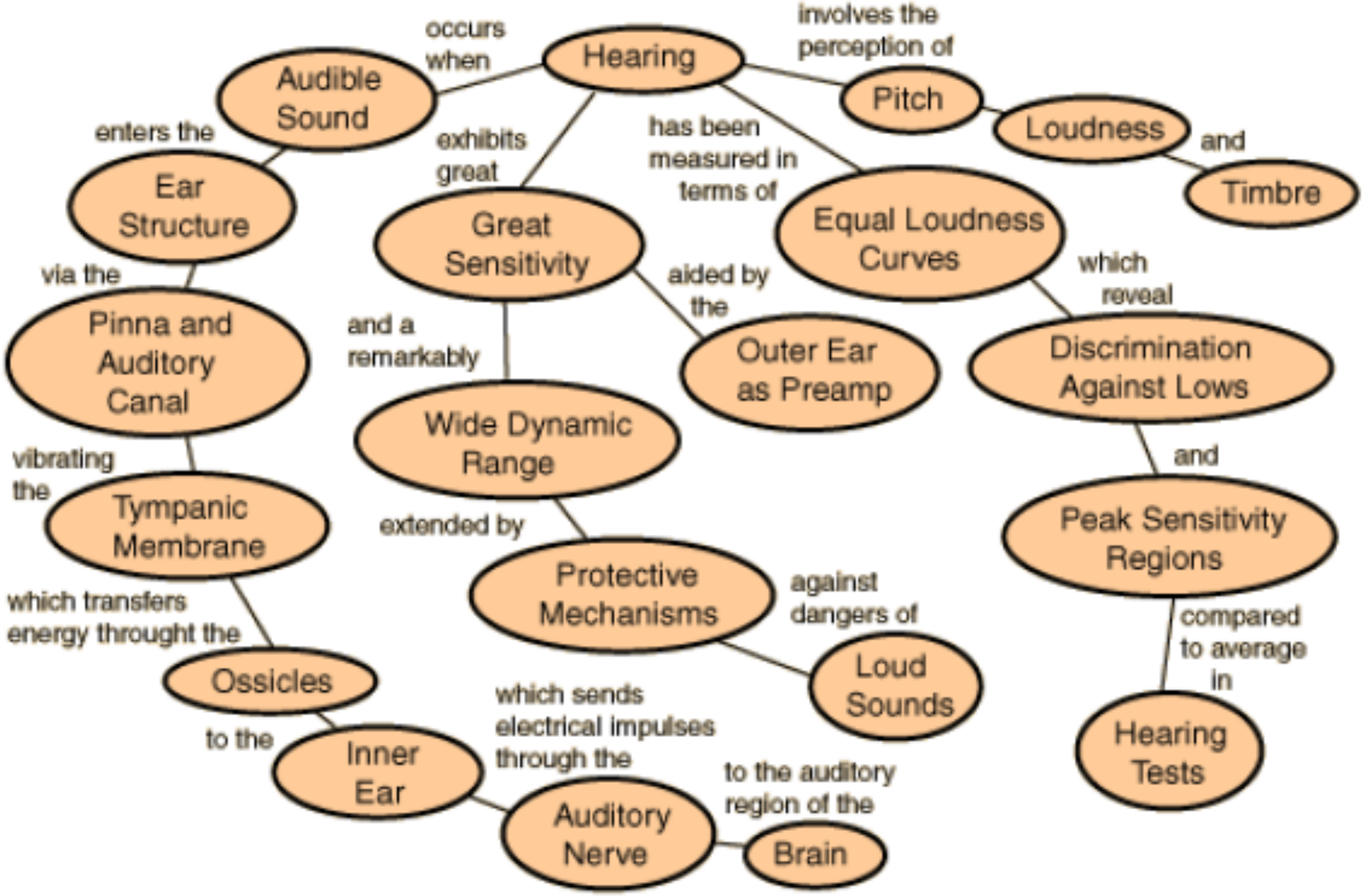
Like light, sound is 'coloured' when reflected, absorbed or passed through objects

Transmission of sound



See http://hyperphysics.phy-astr.gsu.edu/

The Ear



http://hyperphysics.phy-astr.gsu.edu/hbase/sound/hearcon.html

Human Auditory System

Converts mechanical sound energy into neural impulses

Outer ear (Pinna) acts like a radar dish, focusing energy from large area to a small point (auditory canal)

The auditory canal acts as a closed tube resonator: selective freq amplification (~2-5Khz, speaking range)

The middle ear transmits pressure waves to the Cochlea

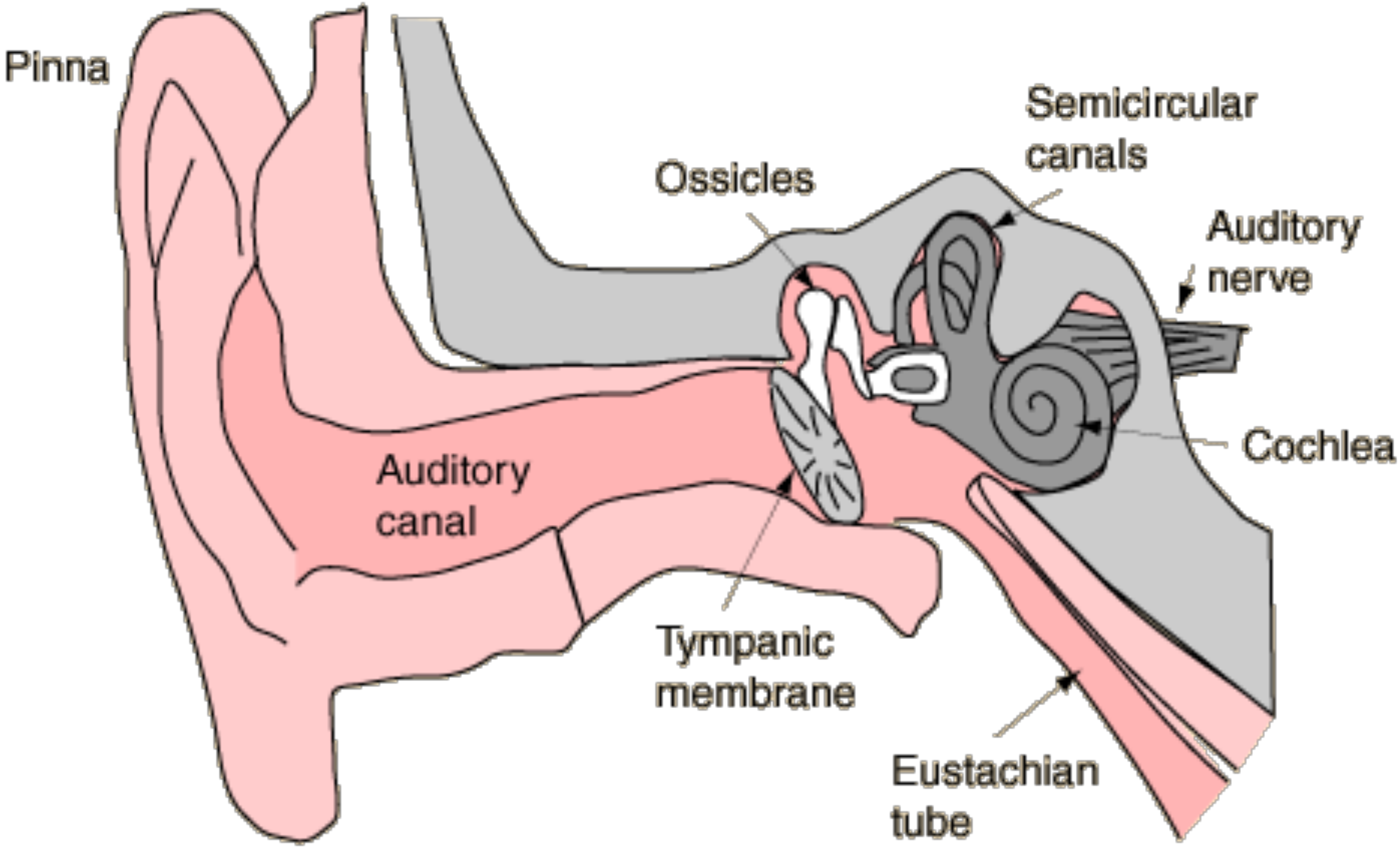
Inside the Cochlea is the Basilar Membrane

- Covered in 15000+ tiny hairs, bent by soundwaves
- These trigger electrochemical impulses

Somehow brain actually converts these to meaningful info

- (various theories)

The Ear

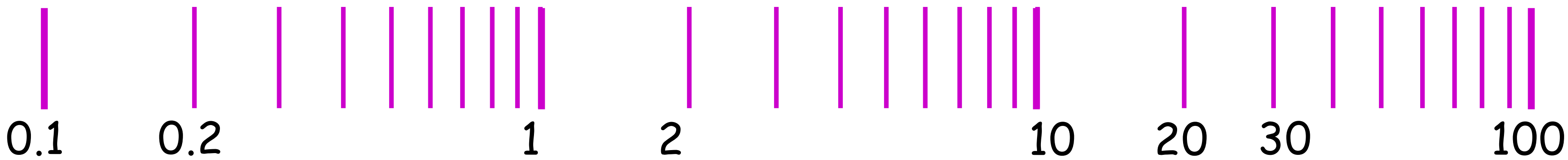

Pinna
Semicircular canals
Ossicles
Auditory nerve
Cochlea
Auditory canal
Tympanic membrane
Eustachian tube

Logarithms and Sound

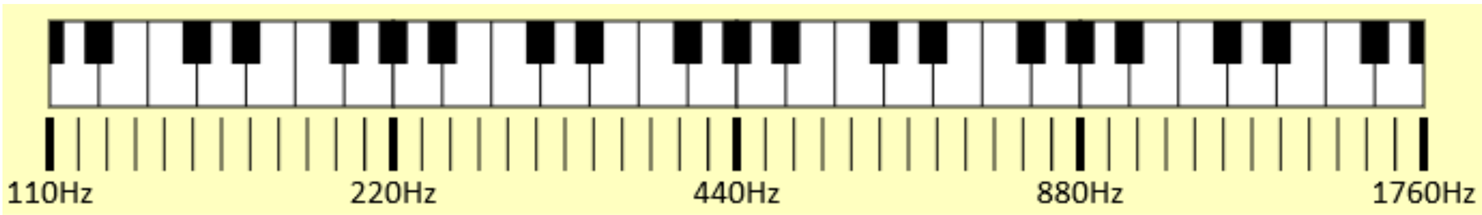
We hear sound over wide ranges of freq : 20Hz to 20kHz

Info from 20..200Hz as important as that 200..Hz

So use log scale … same space when go from f to f*fac



Just like piano keyboard … octave is from f to 2*f



http://www.reading.ac.uk/~shsmchlr/jsfreqresp/AudioSignals.html

Sound power also large range .. 20log10(V) is V in dB

Cybernetics

Auditory Performance

Typical young humans perceive signals 20Hz-20KHz

The band 500Hz-2KHz is particularly sensitive

Human hearing has continuous & complete Field of Regard

(sound can be heard from any direction)

Building blocks of sound perception : pitch and loudness

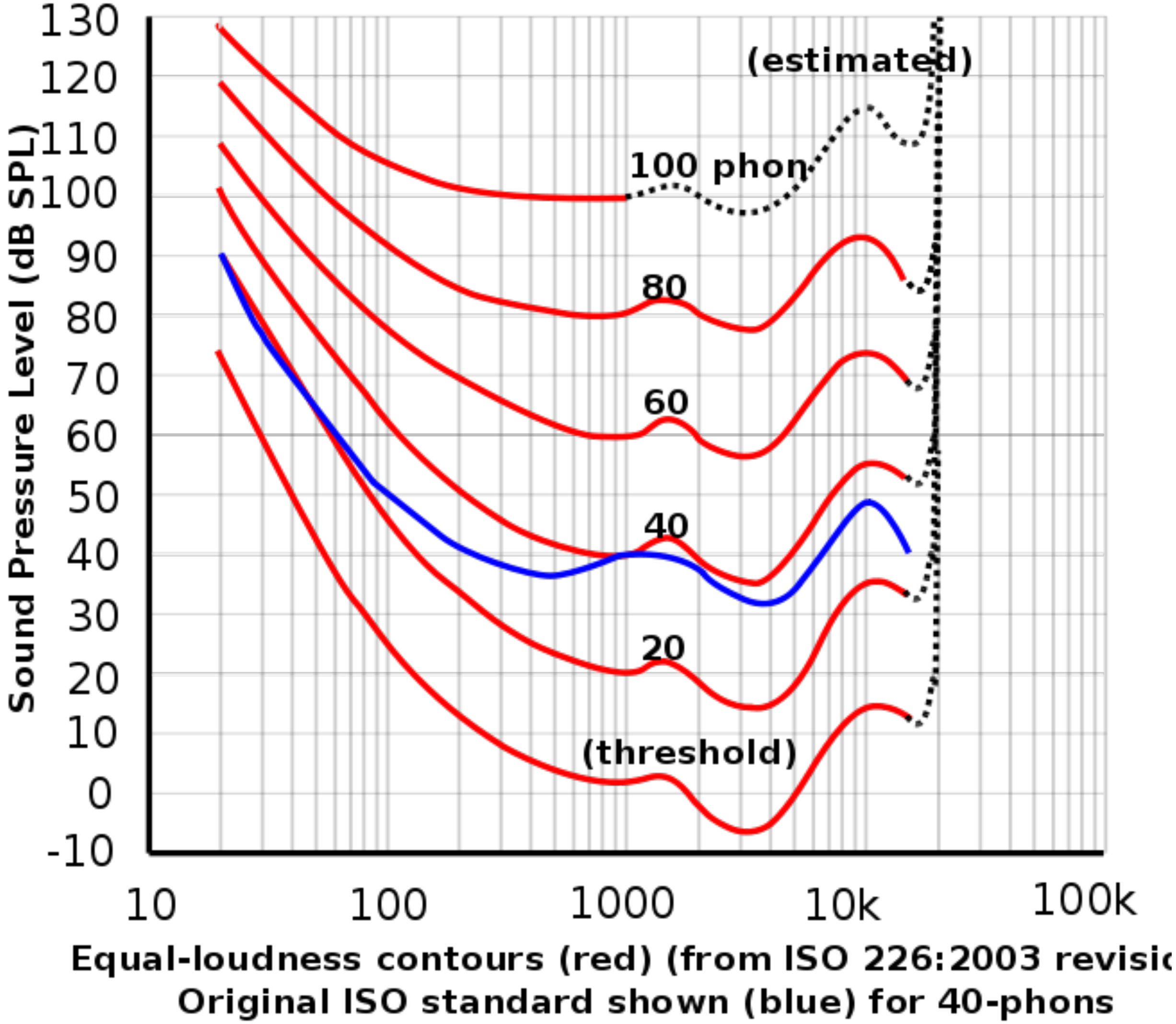
Perception of Pitch ~ logarithmic with frequency

Loudness also logarithmic with amplitude

but affected by other factors including:

duration, previous stimuli, attention and frequency (equal loudness contours)

Equal-loudness_contour



Measure of sound for diff freqs where listener perceives constant loudness (unit phon)

http://en.wikipedia.org/wiki/Equal-loudness_contour

Auditory Perception

Eye has a 2D array (retina), but ears have two points

Yet we distinguish and separate many different simultaneous sounds

'Cocktail Party Effect' – can focus on partic sound even when many other noises

The most powerful cues for this come from grouping sounds by location

Localisation : judging position of sound source in 3D

Requires both direction and distance information.

Spatial Hearing

We locate sound sources based on:

- azimuth cues - position left right
- elevation cues - position up down
- range cues - distance away
- reverberation and echoes
- (and knowledge of the world)

Broad spectrum sounds (many freqs) much easier

carry more info and relate to more auditory cues

Often v. diff to identify direction of single, const freq

Localisation –Time Cues

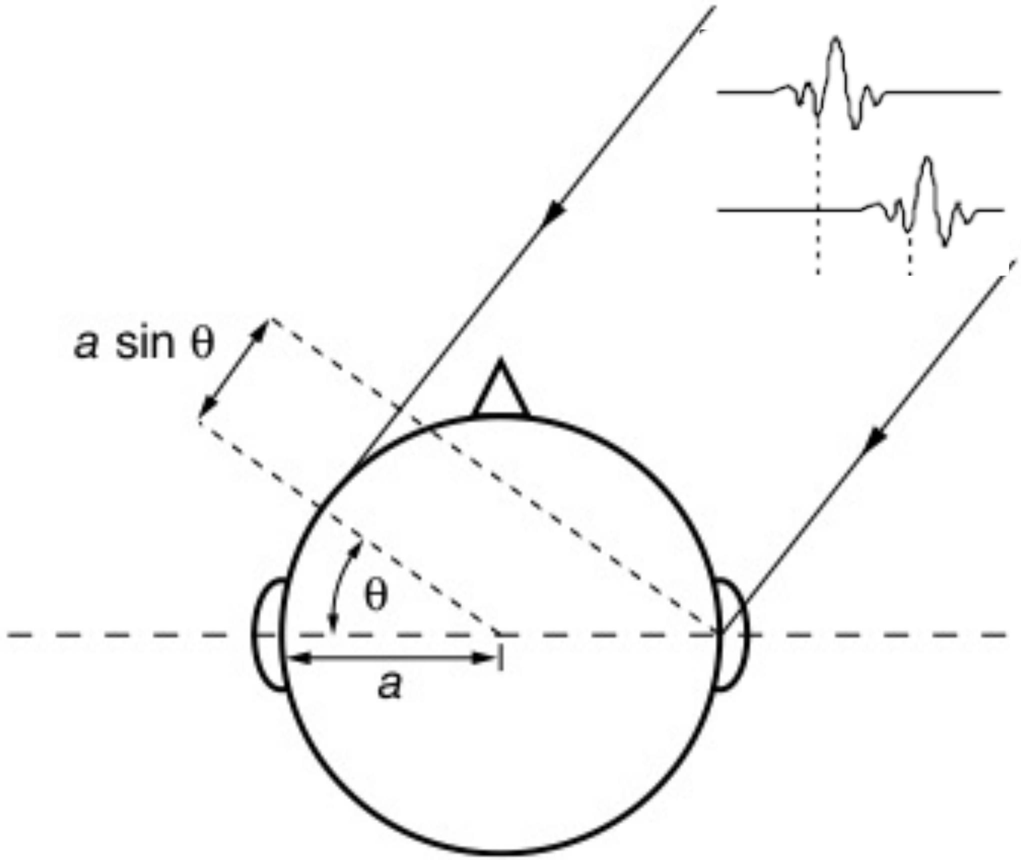
Sound travels at a constant speed.

If not equidistant to both ears, reach one ear first.

Called Inter-aural Time Difference (ITD)

A sound arriving from the side will be delayed by 600-800µs

A human can perceive delays down to 10-20µs



Localisation –Time Cues

For a continuous, non-varying sound the time of arrival at each ear loses meaning

A less powerful cue is Inter-aural Phase Difference (IPD)

The time delay causes a phase change.

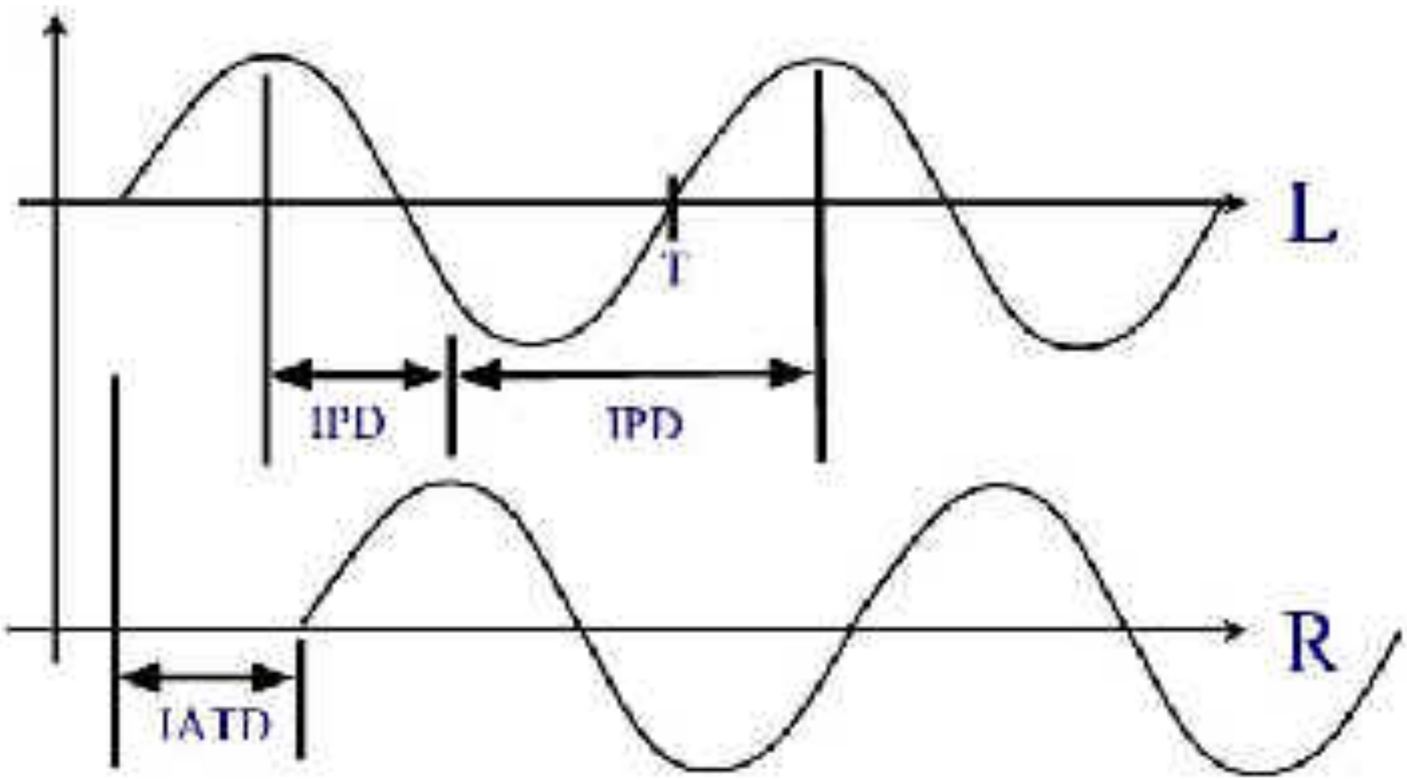
IPD is not consistent, certain angle/frequency combinations will result in no phase shift

At low freq (<1-1.5KHz) , wave length larger than the head

Then neither ITD or IPD work

Hence need one subwoofer in your surround sound system

Showing Phase Difference



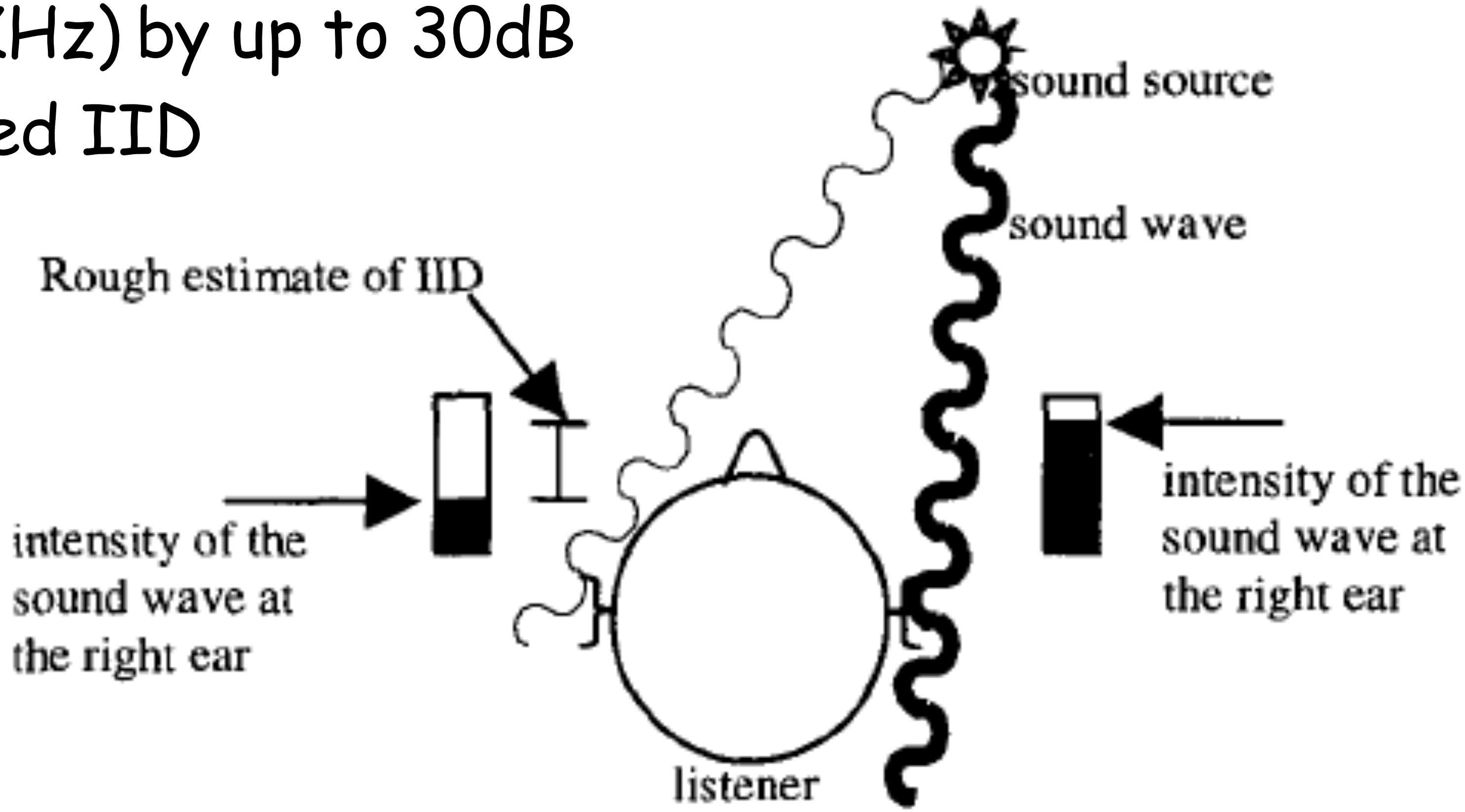
IPD = IATD = T - IATD

Interaural Intensity Difference (IID)

Sounds originating off to one side of the head must pass through the head to reach the far side

This shadowing effect attenuates high frequencies (>4KHz) by up to 30dB

Called IID



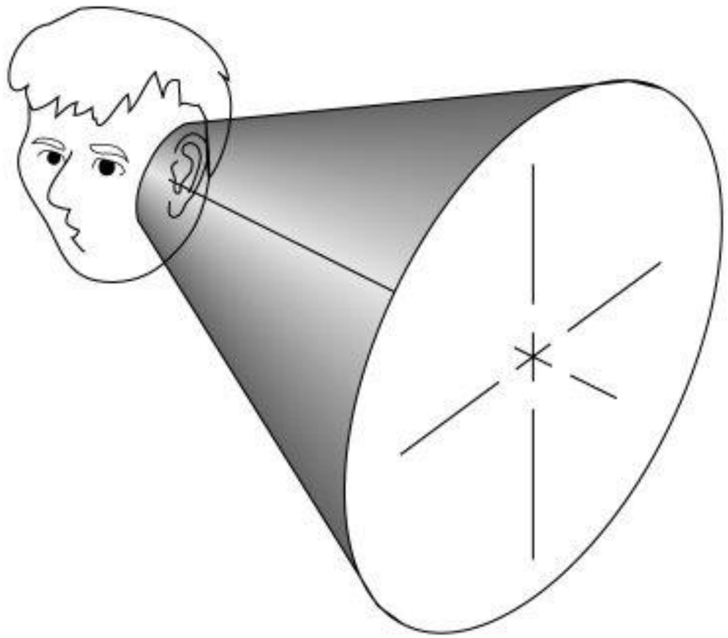
Localisation – Frequency Cues

ITD, IID and IPD do not completely explain the human's sound direction capability

For any point in space there exists a conical set of points equidistant from both ears.

This is the Cone of Confusion

Can be worth tilting your head



Localisation – Frequency Cues

Cone of Confusion : ITD gives azimuth not elevation info

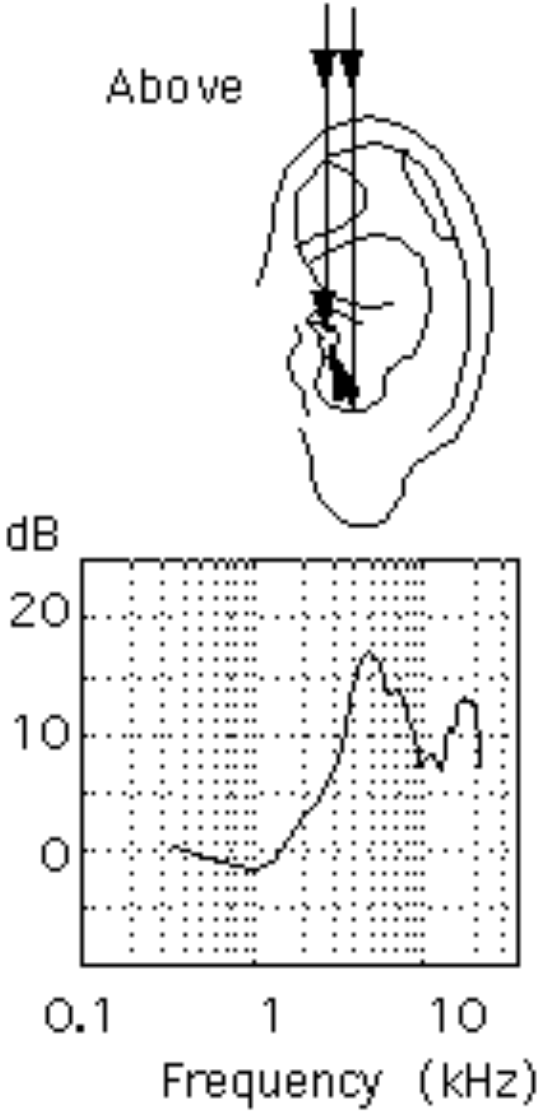
The pinna (outer ear) has non uniform shape

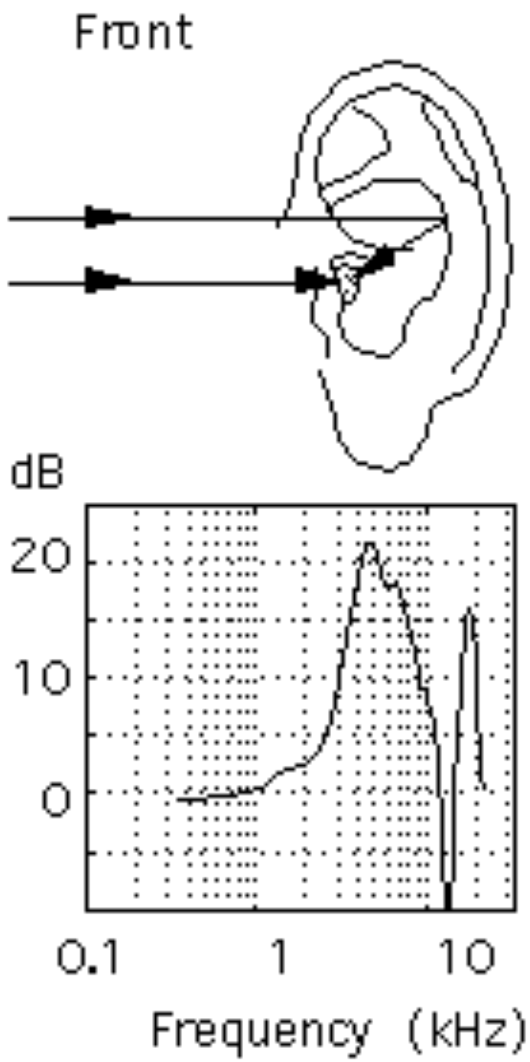
- Makes judgment of elevation possible
- Depending on the direction, different frequencies are attenuated or amplified
- For a sufficiently complex sound this 'colouring' produces strong elevation cues

The interaural distance, head shape and pinna shape make up the Head Related Transfer Function (HRTF)

- Unique to any person
- {Can use shape of pinna for id instead of fingerprints}

Ear better at sounds from side





Localisation – Motion Cues

As with vision, motion parallax v. strong positional cue

If sound of long enough duration

- Listener move head (intentionally or unintentionally)
- Then ITD and IID info change
- Can reduce/remove any ambiguity

In ideal circumstances all these direction cues combine

- for an accuracy of about 1° in azimuth (horizontally) and around 17° in elevation

Localisation – Range Cues

Like vision, depth perception of sound more difficult

Sound intensity reduces with distance

- due to spreading and attenuation in air
- but is not a reliable cue
  - (need lot of knowledge about the source)

Experience of the frequency components of a familiar sound gives an indication of its distance

Again, motion parallax important

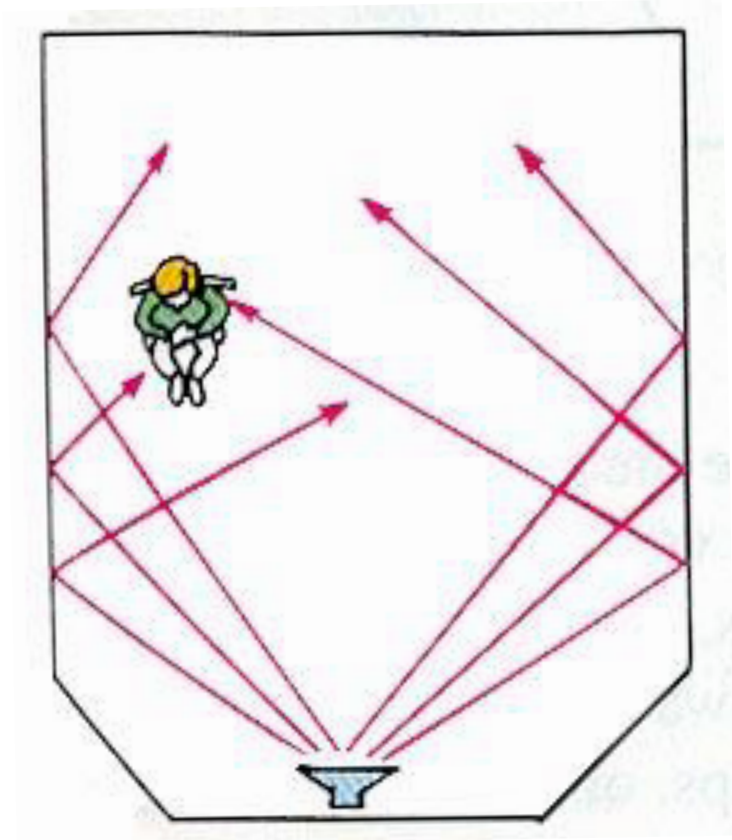
- When head position changes nearby sound sources will change interaural cues more than distant ones

Localisation – Environmental Cues

A human can tell by listening

Whether in a small room or a large auditorium

Whether a sound is coming from round a corner or behind a closed door



How? Brain picks up how sound's freq and time characteristics are modified by the environment

This is primarily due to reflection and occlusion

Localisation – Environmental Cues

Reflections (reverberation) are a function of time

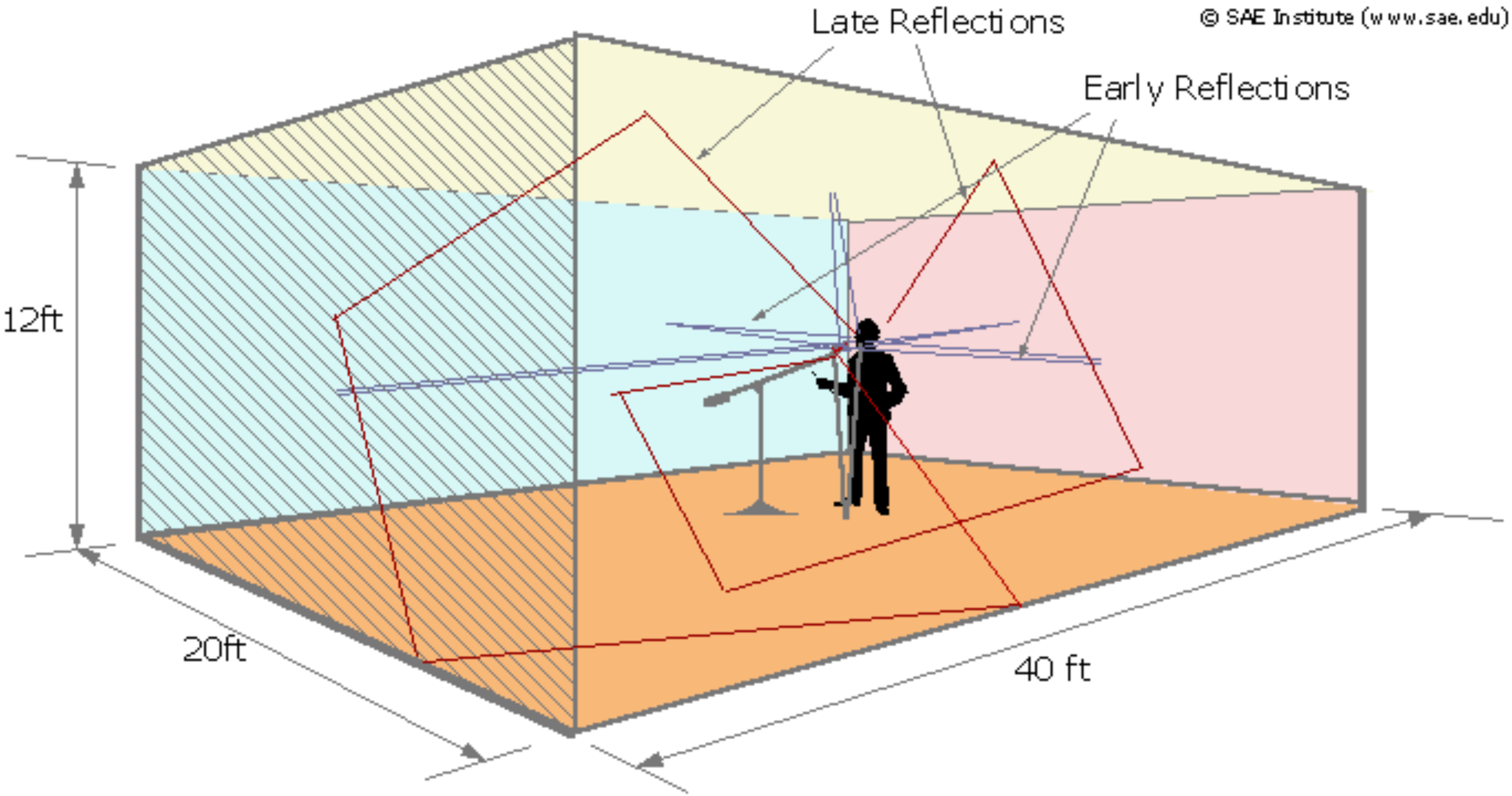
The complete audio picture heard by a listener comprises

- The direct path sound
- First/early reflections bounced off a obstacle
- Secondary/Late reflections travelled further

Reflections also affected by absorption and diffusion characteristics of the materials they encounter

- These alter the frequency spectrum

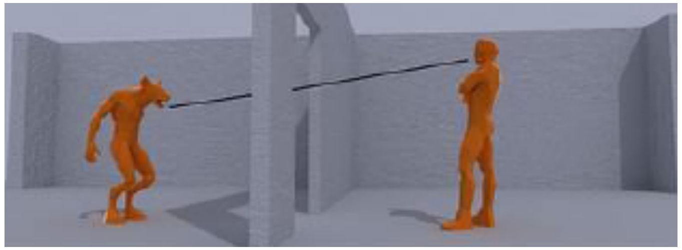
Image



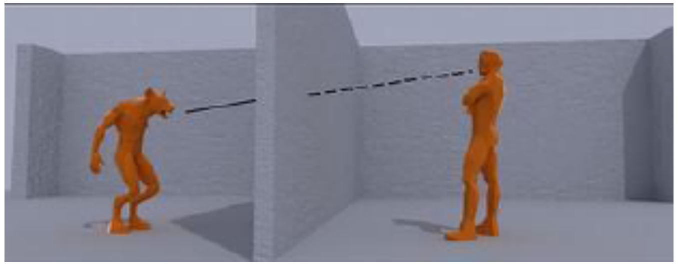
http://www.sae.edu/reference_material/audio/pages/Reverb.htm

Localisation – Environmental Cues

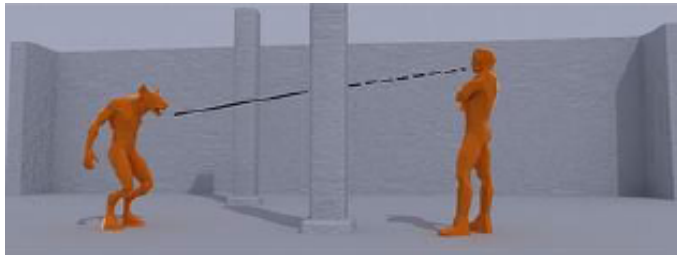
Occlusions are primarily a function of frequency
When an obstacle blocks a direct sound path or one of its reflections higher freq components are attenuated



Exclusions
Direct Path clear
Reflections muffled



Occlusions
Direct Path muffled
Reflections muffled



Obstructions
Direct Path muffled
Reflections clear

Cybernetics

Spatial Hearing – Knowledge of World

Expectation and familiarity

we perceive sounds to come from places where we expect sounds to come from

Visual cues

we perceive sounds to come from sources that look like they're making sounds

(exploited by ventriloquists)

Given what we have heard, we now can consider how to produce sound for virtual worlds..

Audio Systems

Sounds are transformed by environment and listener (HRTF) before reaching inner ear

- Former tells listener about position of source in environment and some properties of the environment
- From HRTF the listener receives info about the sound source's position relative to their head

For VR, 3D audio processing techniques needed so sound

- appears in both correct position and environment type

How these transforms are implemented

- closely related to the display type,
- two categories: headphones and loudspeakers

Stereophonic Headphones &HRTFs

The perfect 3D audio system can make a sound appear to originate from any direction

Need manipulation of cues humans have learnt for location

One method : provide unique audio image to each ear as close as poss to ear canal

headphones

Then immediate environmental effects are removed

Easy to create an artificial soundscape

But, also removed are the cues produced by the HRTF

Stereophonic Headphones and HRTFs

Recreating interaural time differences is not difficult

- a sound is played in each ear at different times

Recreating the effects of the pinna is much more difficult

- Due to the required complexity and high quality of the real time frequency transforms
- And the fact that everyone's ears are different

An unrealistic HRTF can result in

- high front-back confusion and internalisation
  - (sounds appear very close to or inside the head)

Making an HRTF

A HRTF can be found in two ways:

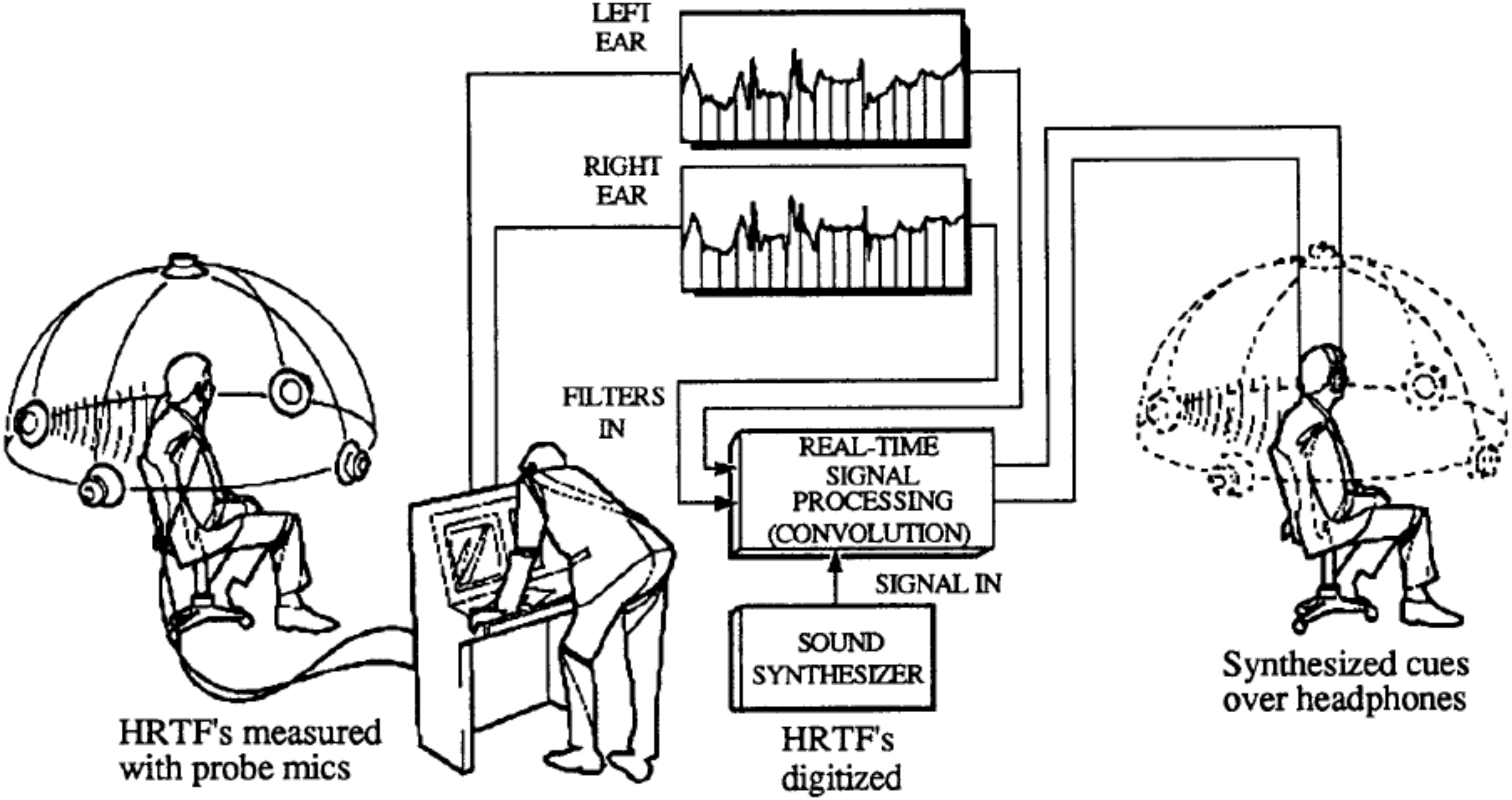
- Placing small microphones in the ear canal and recording the transformation between a tone played at a known location and the recording from within the ear
- Taking measurements of the pinnae and calculating acoustic models

Even though the HRTF should be personalised to get good results, this is not practical for general purpose VR

Instead, averaged transforms from multiple humans or dummy heads are used

Full HRTFs can also result in distorted sound so simplified versions are often used, even less realism

Measuring the HRTF



On Head tracking for VR

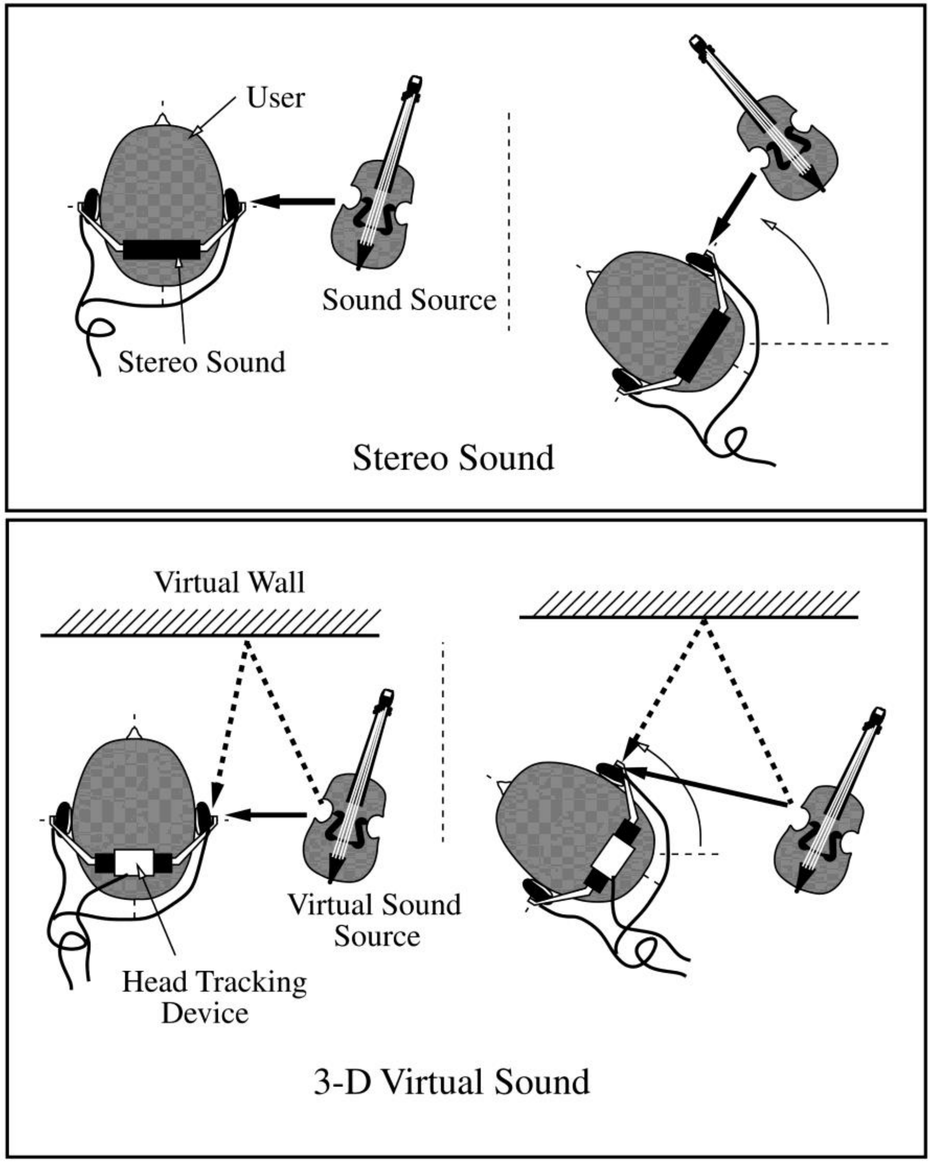
Parallax effect very strong cue for auditory localisation.

Head tracking and updating of listener's position important

For desktop displays need less: user's head in same direction

For HMDs and immersive environments, any absence of head tracking

- very noticeable,
- reducing presence
- may cause disorientation.



Loudspeaker & Multichannel Displays

Alternative to headphones for recreating the full HRTF

- Place a number of loudspeakers around the user,
- utilising at least some of the natural direction cues

Creating a sound which appears to originate from the direction of a physical speaker is very easy

More diff. to create a sound which moves around user

- does not originate exactly at a physical speaker,
- a so called 'Phantom Source'

Many techniques for loudspeakers recreating spatial cues

- Cross-talk cancellation, Vector Base Amplitude Panning, Ambisonics and Wave Field Synthesis

Vector Based Amplitude Panning

Virtual source shares the gain between the nearest speakers to the direction of actual source of user

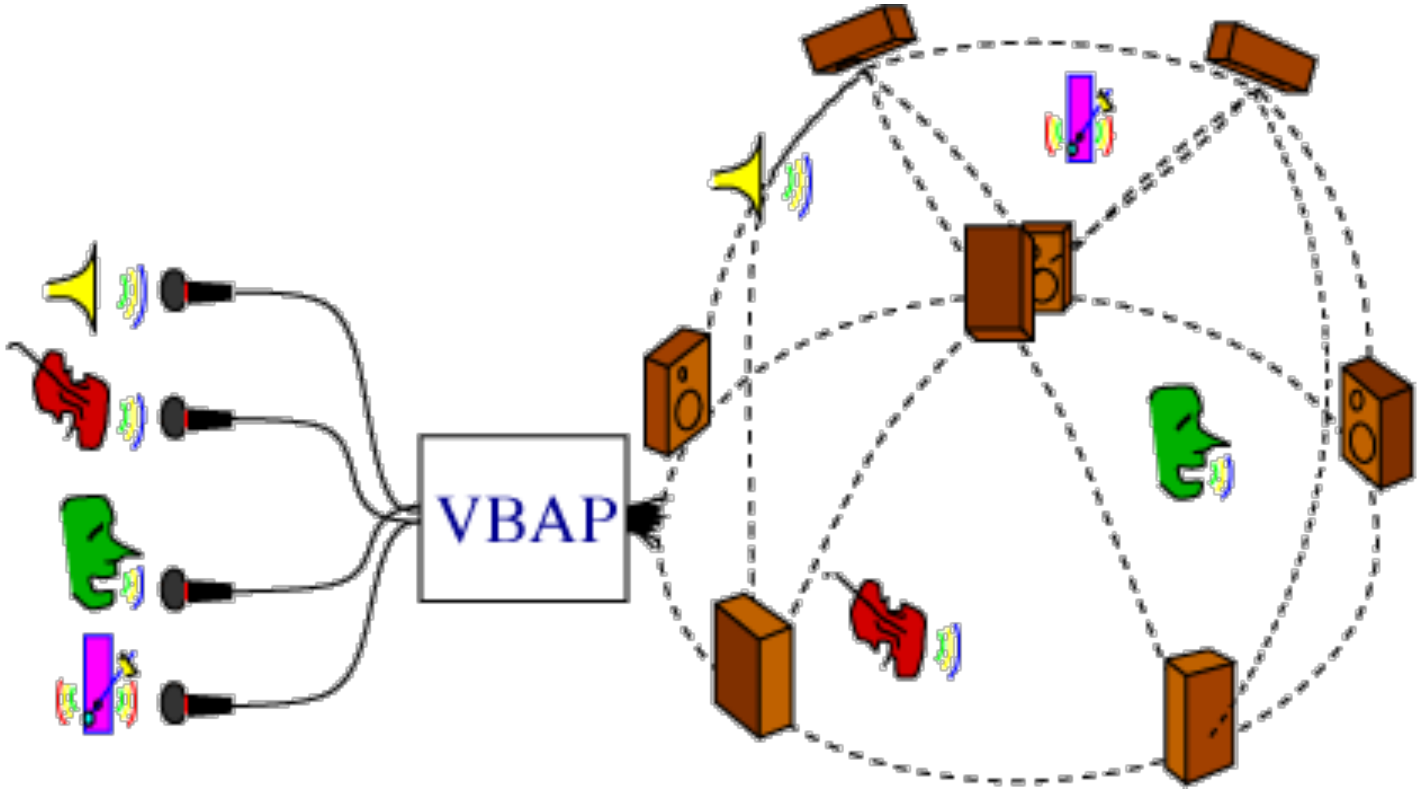
- If a source lies in exactly the same direction as a speaker all the gain will be channelled through it
- If the source lies evenly between two speakers half the gain will go to each

Signal well defined as minimum number of speakers used

It is difficult to generate a virtual sound which is perceived very far from a real speaker

Effectiveness related to density /direction of loudspeakers

Figure



http://legacy.spa.aalto.fi/research/cat/vbap/

V. Pulkki "Compensating displacement of amplitude-panned virtual sources." Audio Engineering Society 22th Int. Conf. on Virtual, Synthetic and Entertainment Audio pp. 186-195. 2002 Espoo, Finland

Cross-Talk Cancellation

With VBAP, Virtual sources appear close to real speakers

Because there are many cues, (ITD and pinna)

These indicate to the listener the sound's origin

Difficult to recreate these cues as both ears receive the output of every speaker - known as cross-talk

If know approx position of listener relative to speakers

Can do cross-talk cancellation

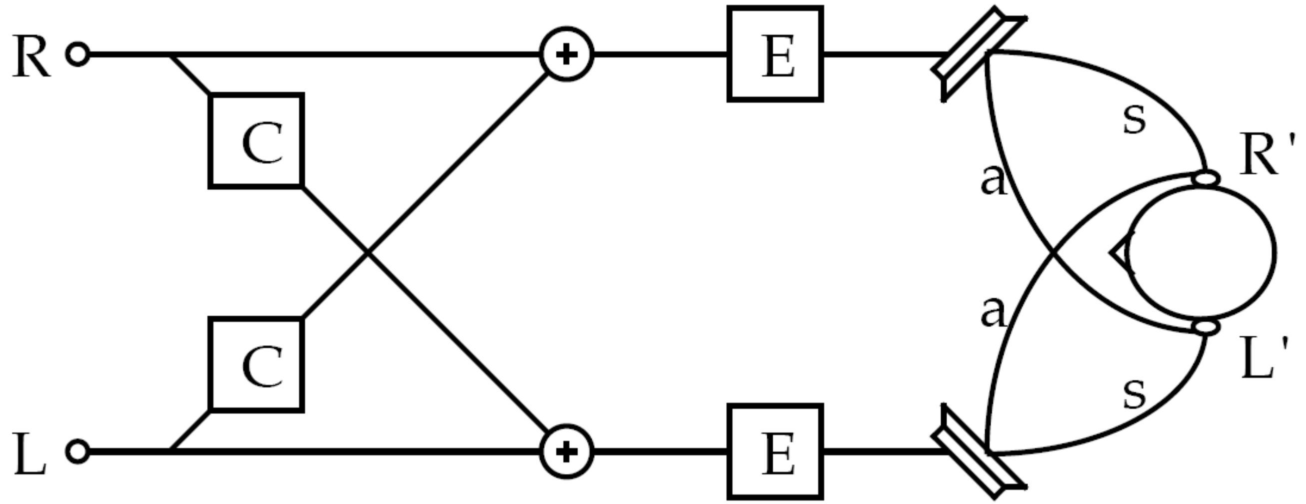
To remove the unwanted signal from one speaker,

add signal 180^0 out of phase to that of diff speaker

delayed so both signals arrive at ear at same time

Cancellation sets parts of each speaker's output ear gets

Cross-Talk Cancellation



To create a virtual source, still need artificial HRTF, so limitations described for headphones apply equally here.

Many consumer 3D audio systems use cross-talk cancellation with HRTFs for two speaker configurations but something more like VBAP for rear speakers

Ambisonics

Ambisonics encodes what the user should hear as

a sound pressure (W) and

a set of direction vectors (X, Y, Z) of particle velocity

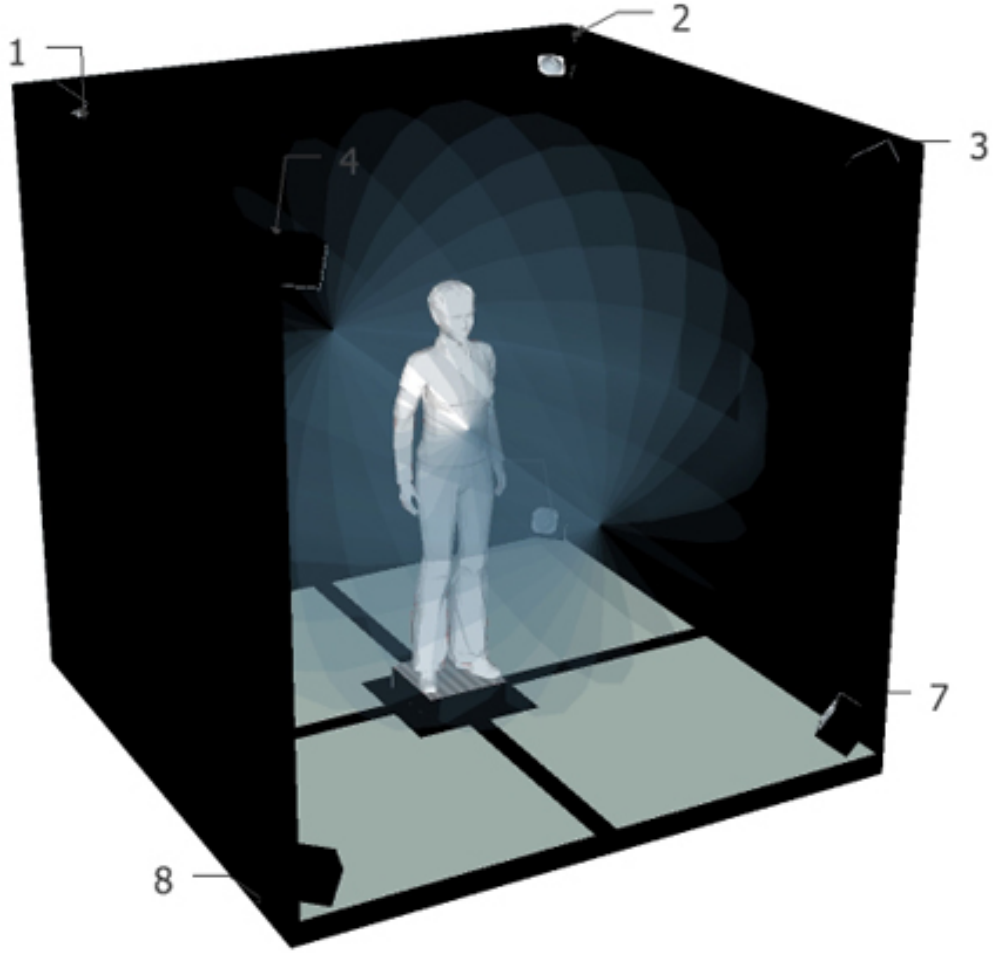
In simplest decoding scheme the output of each speaker is linear combination of the four channels dependant on the position of the speaker in relation to the user

Most implementations use 8 speakers at corners of a cube

Generally all speakers are used to localise a sound in any direction

As such spatial definition can be quite low in some cases

Eight Speakers



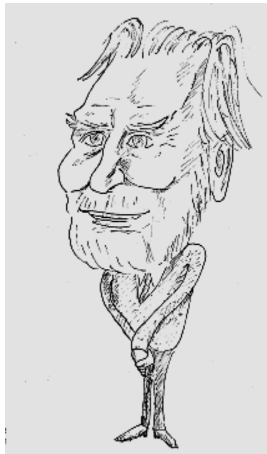
Ambisonics

Unusually the signal is the same irrespective of the number of speakers or where they are

Has picked up some support in computer games (see wikipedia for a list)

Good BBC interview here: http://www.ambisonic.net/

University of Reading link: Professor Peter Fellgett was part of the group who originally developed the theoretical framework for Ambisonics



Wave Field Synthesis, WFS

WFS based on principle that any wave-front can be regarded as a superposition of elementary spherical waves

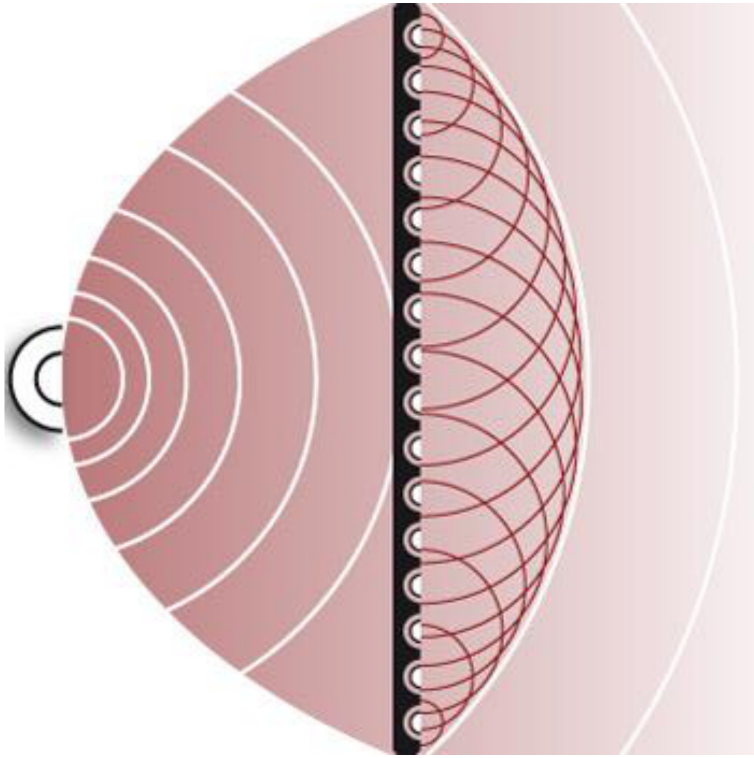
The ideal implementation is an anechoic room covered with a large number of individually driven speakers

Each speaker reproduces the portion of a wave front that, if originating from a specific source, would be passing through it at any point in time

Theoretically a highly realistic sound field could be created, but a number of practical restrictions exist:

- Reflection free (anechoic) surfaces
- The density of speakers

Synthesising a wave front



Multiple speakers synthesising a wave front
http://www.iosono-sound.com/technology/wave-field-synthesis

Loudspeakers Vs Headphones

Headphones:

- Eliminate actual environmental effects
- Lose actual HRTFs
- Custom HRTFs are inaccurate and can distort sound
- Difficult to get generic HTRFs but time consuming/awkward to generate specific
- Sounds seem closer (internalised)
- Directionality is good

Loudspeakers Vs Headphones

Loudspeakers:

- Have to deal with actual environment
- Get real HRTF up to a point
- Directionality from physical speakers is very good, i.e. sounds from behind very realistic
- Really need speakers from every direction
- If want true phantom sources need to isolate with CC filters
- Important part of localisation is parallax, get this well

Some links

Virtual Barber Shop:
http://www.youtube.com/watch?v=IUDTlvagjJA

http://www.noogenesis.com/binaural/binaural.html

http://www.ambisonic.net/

http://ixbtlabs.com/articles2/sound-technology/

http://www.soundonsound.com/sos/oct01/articles/surroundsound3.asp

Next time we consider feeling a virtual world …

Haptic Perception

What is Haptics?

- Pronunciation: \hap-tik\
- Function: adjective
- Etymology: International Scientific Vocabulary, from Greek *haptesthai* to touch
- 1 : relating to or based on the sense of touch
- 2 : characterized by a predilection for the sense of touch <a *haptic* person>
- www.merriam-webster.com

Cybernetics

What is Haptics

"Haptics is now commonly viewed as a perceptual system, mediated by two afferent subsystems, cutaneous and kinesthetic, that most typically involves active manual exploration. Whereas vision and audition are recognized for providing highly precise spatial and temporal information, respectively, the haptic system is especially effective at processing the material characteristics of surfaces and objects."

The haptic system uses sensory information derived from mechanoreceptors and thermoreceptors embedded in the skin ("cutaneous" inputs) together with mechanoreceptors embedded in muscles, tendons, and joints ("kinesthetic" inputs). (Lederman & Klatzky, 2009).

Touch & Haptics

Touch describes sensations produced when objects come into physical contact with the body.

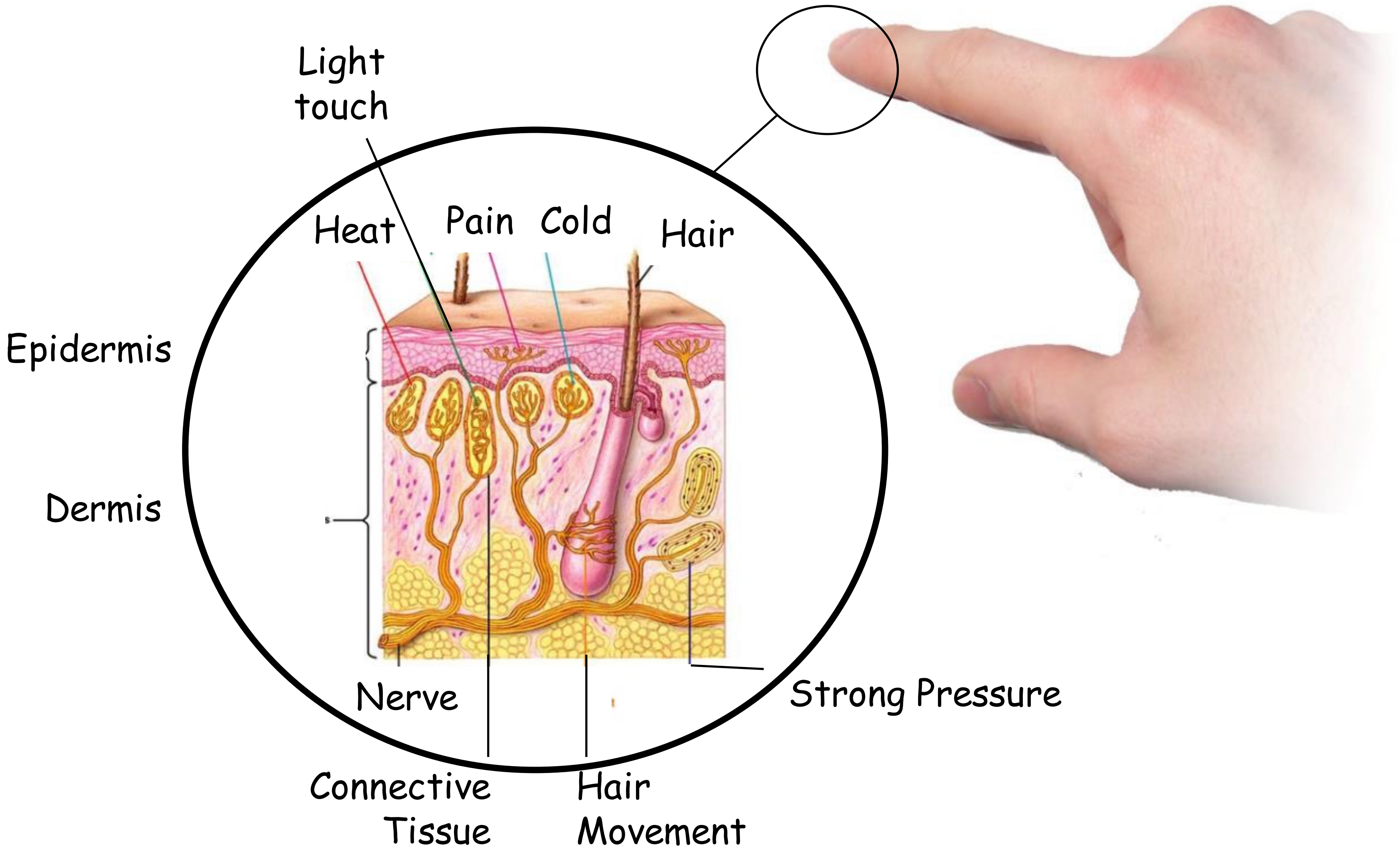
Tactile sensation originates in the skin

Small scale surface geometry, pressure, texture, vibration, temperature

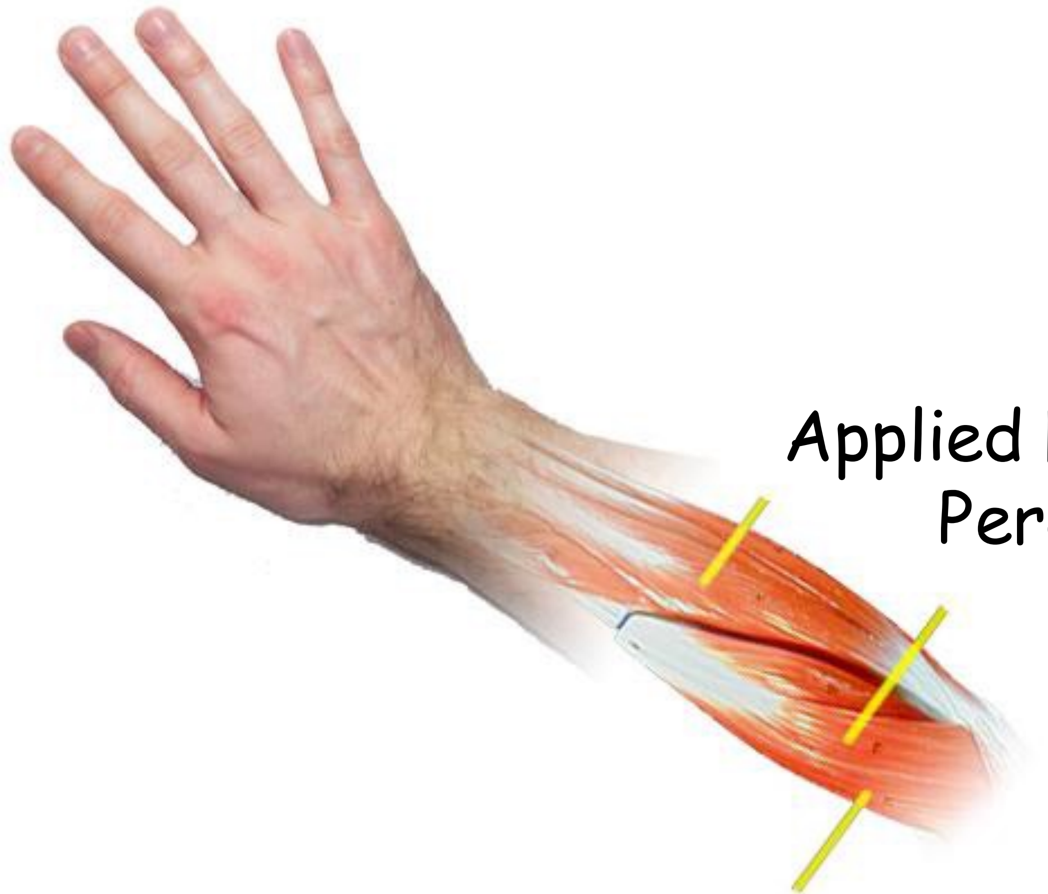
Kinaesthetic/Proprioceptive sensation occurs in the muscles, joints and tendons

Weight, force, stiffness, large scale shape

Tactile Sensation



Kinaesthetic …



What is Haptics?

Haptics includes tactile and kinaesthetic sensation

In Psychology & VR haptics usually refers to Active Touch

For Active Touch to occur changes in touch sensation must correlate with purposeful movement

- Using a mouse isn't usually considered haptics even though you can feel it and sense its position
- This is because there is no feedback

A haptic device gives touch feedback based on input

Physiology of Haptics - Tactile

Relies on sensors in and close to the skin

Three types:

- Thermoreceptor (Temperature)
  - Measure RATE at which skin temperature changes.
  - Two kinds, warm and cold.
- Nociceptor (Pain)
  - Don't generally include pain sensation in haptics as is a subjective experience (not referenced to world)
  - Usually better to know how to avoid than recreate!
- Mechanoreceptor (Deformation and Vibration)

Physiology of Haptics - Tactile

Average person has approximately 2m^2 of skin which contains a variety of different receptors

As with the retina in the eye, distribution of receptors is not even.

Denser populations occur where high touch acuity is required (finger tips, lips)

The Sensory Hommunculus represents the proportion of brain area devoted to different parts of the skin

Types of Mechanoreceptor

Glabrous (non-hairy) area of the hand (palm and finger tips) contains four types of mechanoreceptor

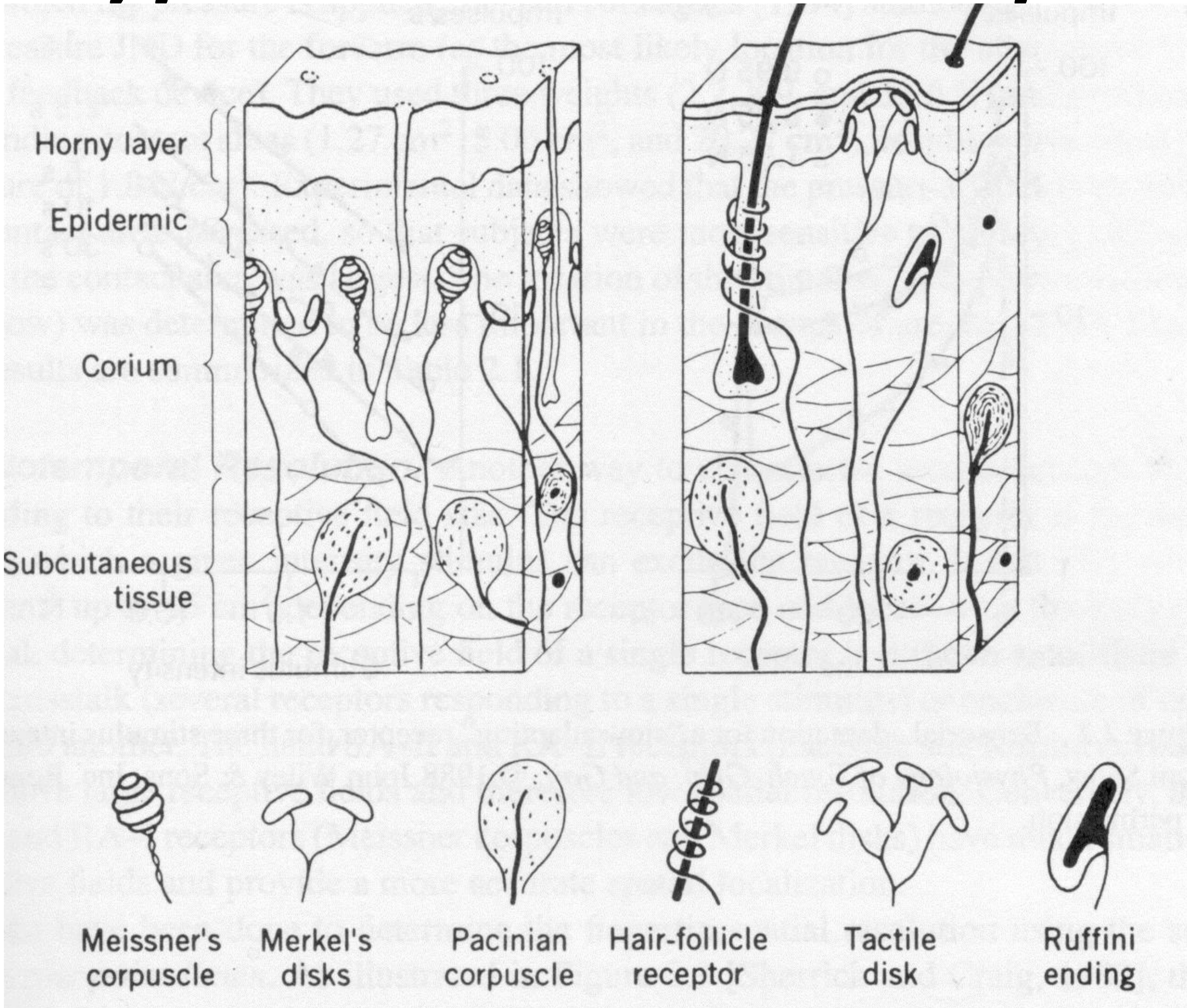
40% are Meissner's corpuscles – detect movement across the skin; light touch

25% are Merkel's disks – sustained touch and pressure

13% are Pacinian corpuscles – deeper in skin (dermis); acceleration sensors; most sensitive to vibrations of about 250 Hz

19% are Ruffini corpuscles – detect pressure, skin shear, lateral skin stretch

Types of Mechanoreceptor



Properties of Mechanoreceptors

Sensorial Adaption : decrease in electrical response from a receptor over time (for a constant stimulus)

- Slowly Adapting (SA) receptors keep firing for a while
- Rapidly Adapting (RA) quickly forget and output rapidly decreases

Comparison of Mechanoreceptors

Table 3.3 Comparison of various skin mechanoreceptorss

Receptor Type	Rate of Adaptation	Stimulus frequency (Hz)	Receptive Field	Function
Merkel Disks	SA-I	0–10	Small, well defined	Edges, intensity
Ruffini Corpuscles	SA-II	0–10	Large, indistinct	Static force, skin stretch
Meissner Corpuscles	FA-I	20–50	Small, well defined	Velocity, edges
Pacinian Corpuscles	FA-II	100–300	Large, indistinct	Acceleration, vibration

Based on Seow [1988], Cholewiak and Collins [1991], and Kalawsky [1993]

Spatial Resolution

Density and receptive field size affects spatial resolution

Receptive field size : area of skin which can excite a receptor

If the sensor has a large receptive field – it has low spatial resolution

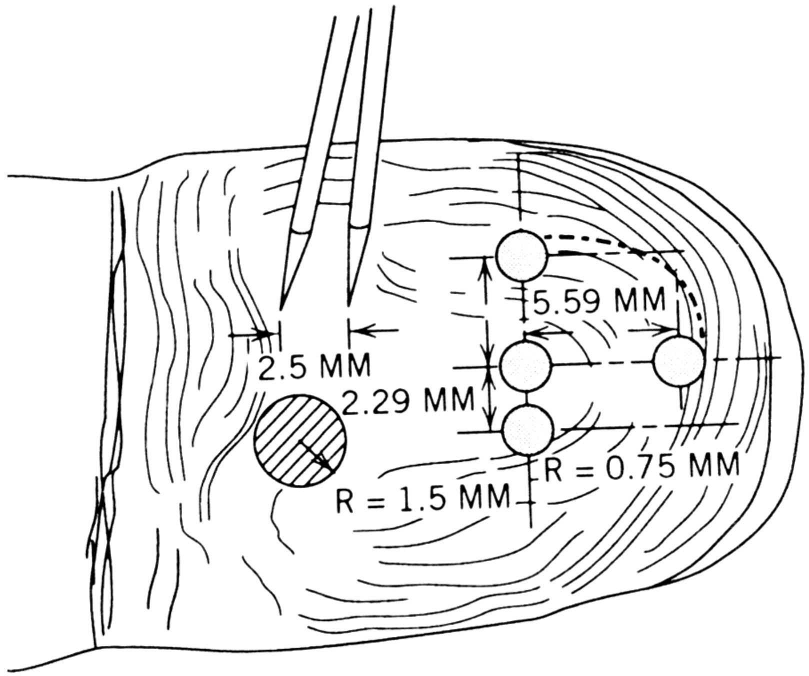
If the receptive field is small, has high spatial resolution

Varies depending on receptor type and location on the body

Spatial Resolution

Acuity is measured using the two-point threshold test

- Minimum separation between two points of contact that can be achieved before they are perceived as one
- From 2mm on the finger pad, 1mm on the palm to 30mm on the forearm and up to 70mm on the back.



Localisation : ability to judge distance between two temporarily separate stimuli … accuracy up to 0.17mm

Proprioception & Kinaesthesia

Proprioception is (unconscious) awareness of positions and movements of the body and the amount of force being applied by the muscles from internal mechanisms

Kinaesthesia is often used interchangeably with proprioception but places a stronger emphasis on motion and force and is not purely internal (includes vision)

No distinct line between Proprioception and tactile cues:

Judgment of properties such as stiffness come from both.

Without knowledge of limb position and velocity, the output of the mechanoreceptors becomes a jumble of disconnected sensations.

Proprioception & Kinaesthesia

The sensors for proprioception are in the muscles, tendons and joints

Many of the receptors are the same as those found in the skin such as Ruffini Endings and Pacinian Corpuscles

Give information about joint angles and velocities

Also stretch sensitive receptors connected between tendons and muscles called Golgi Organs

Gogli Organs also activate a 'cut-out' system which protects the muscles from applying excessive damaging force.

Proprioception & Kinaesthesia

Proprioception has open-loop elements and includes commanded forces in internal model

Measurement of positions, forces and angles is not absolute and can drift. (Floating arms illusion)

Accuracy of angular position decreases with joint size and limb length - about 0.2° at the hip to 0.7° in the metatarsophalangeal joint

The open-loop nature of proprioception can be exploited in touch interface design

Psychology of Haptics

'The frequently made suggestions that vision is "more accurate than" or "dominates" touch seem to miss the point. Suggestions of this kind derive from the notion that touch exists only to do what vision can do better. Touch is not simply an inferior form of vision, nor hearing. Touch, as touch, has its own capabilities and limitations.'

- S. J. Lederman

Psychology of Haptics

Touch is uniquely different from vision and hearing

- Many secondary qualities unique to touch (rough, heavy, cold etc.)
- Perception of sight and sound occurs at a distance, touch requires direct contact
- Touch is a two way process, touch can alter the environment as well as receiving information about it

Form Perception

As with visual shape perception, touch based shape perception is predominantly concerned with finding boundaries

but language is in terms of forces and positions.

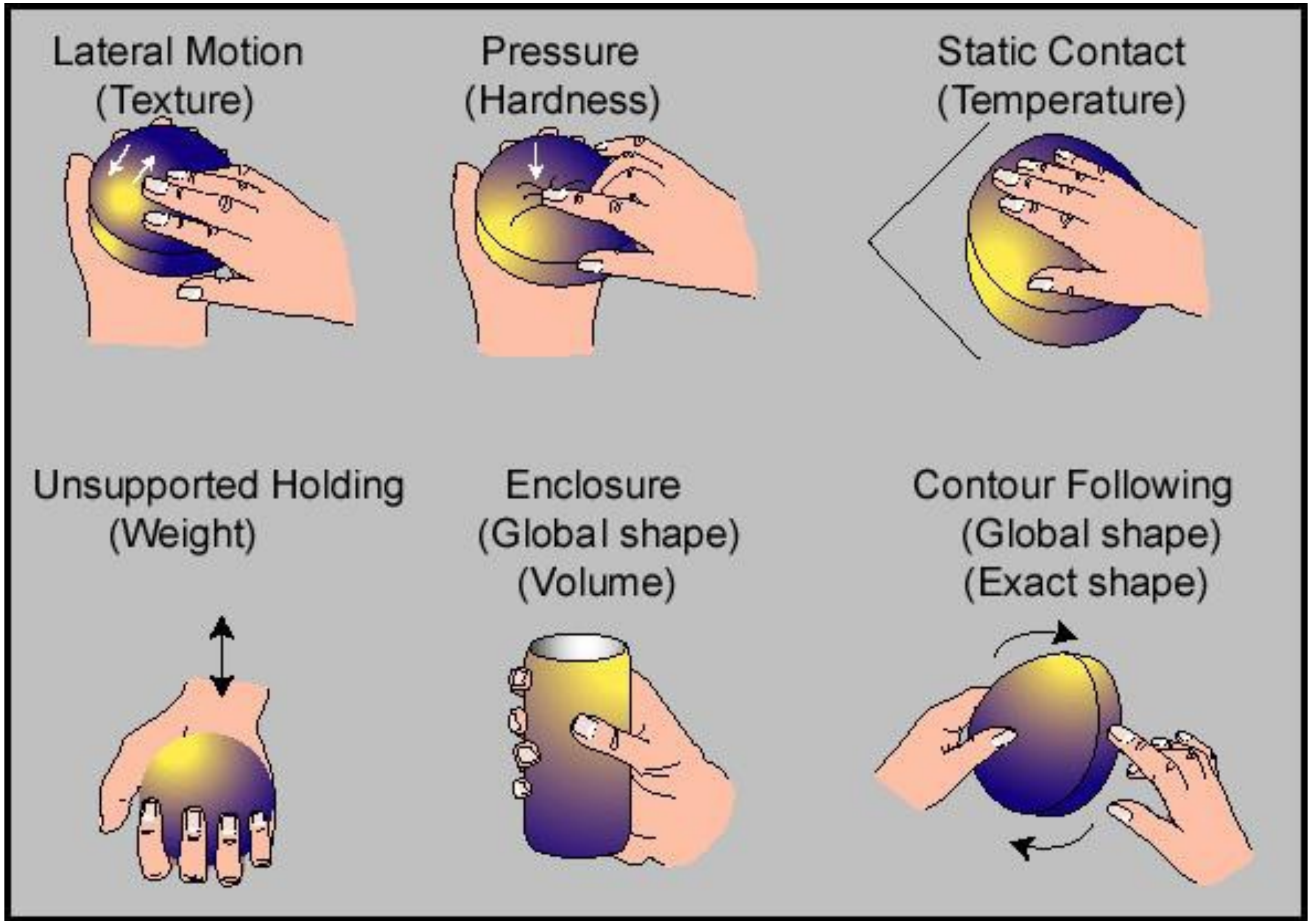
The mechanoreceptors give small scale shape information in terms of lines, edges, and corners etc. which occur as a result of forces deforming certain areas of the skin.

The proprioceptive system gives shape on a larger scale by measuring force direction on joints and mapping them to positions in space

Exploratory Procedures

Humans use consistent and predictable movements when haptically investigating an object

The so called Exploratory procedures (EPs)



Texture Perception

Highly complex process, driven by tactile sensors

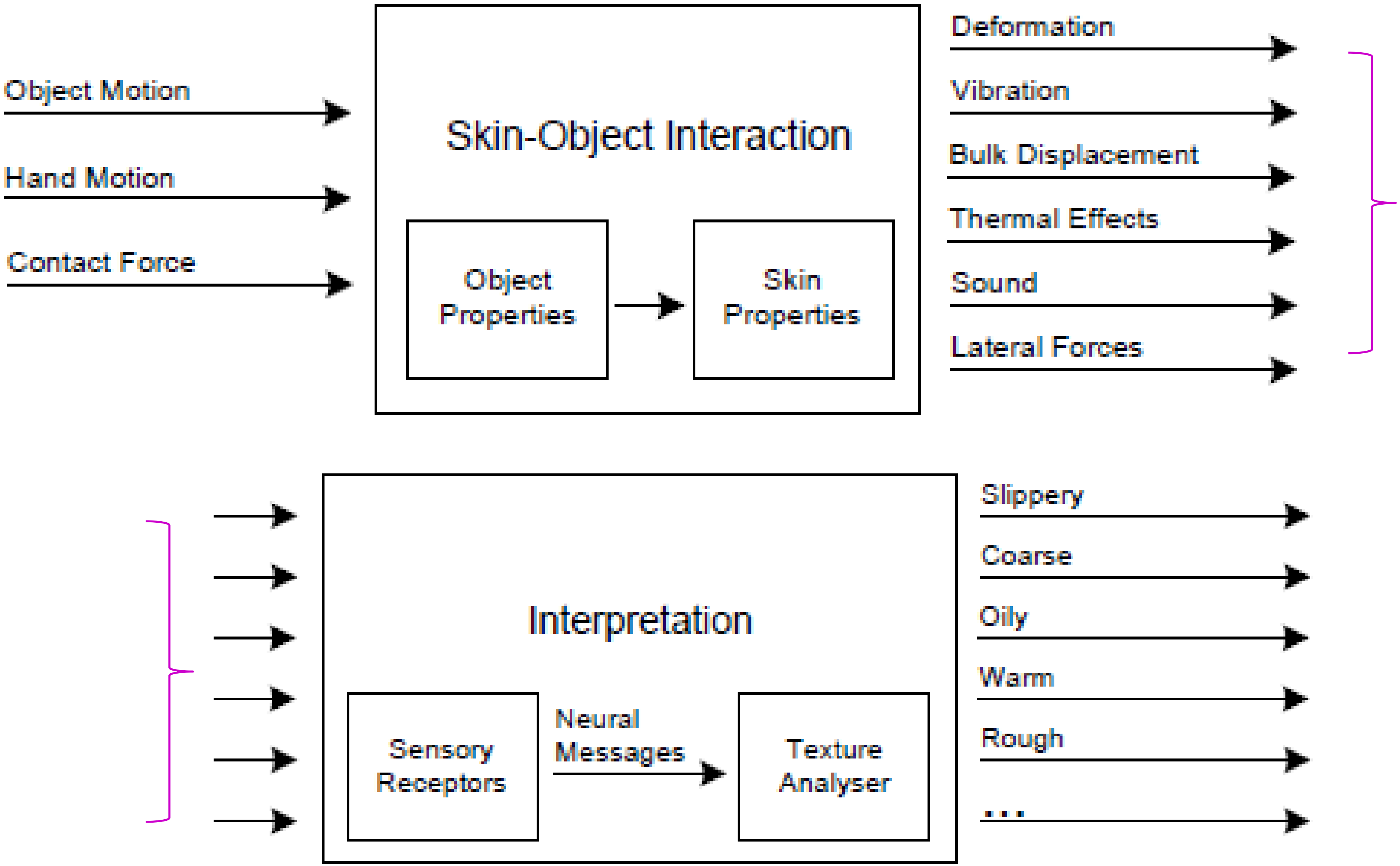
Different sensory channels are fused together to create the experience of texture

The complex nature of the process gives rise to a diverse set of perceptual experiences:

- rough, sharp, cold, sticky, slippery, rubbery, oily and grainy etc.

Not only the object's properties affect the output, skin properties such as dryness also have an effect.

- For example, cold skin makes surfaces seem smoother as the skin's elasticity is reduced



Cybernetics

Haptic Display Technologies

We have explored aspects of touch

- How humans feel the real world

Next we consider the technology available

- for feeling virtual worlds

Haptic Displays - Tactile

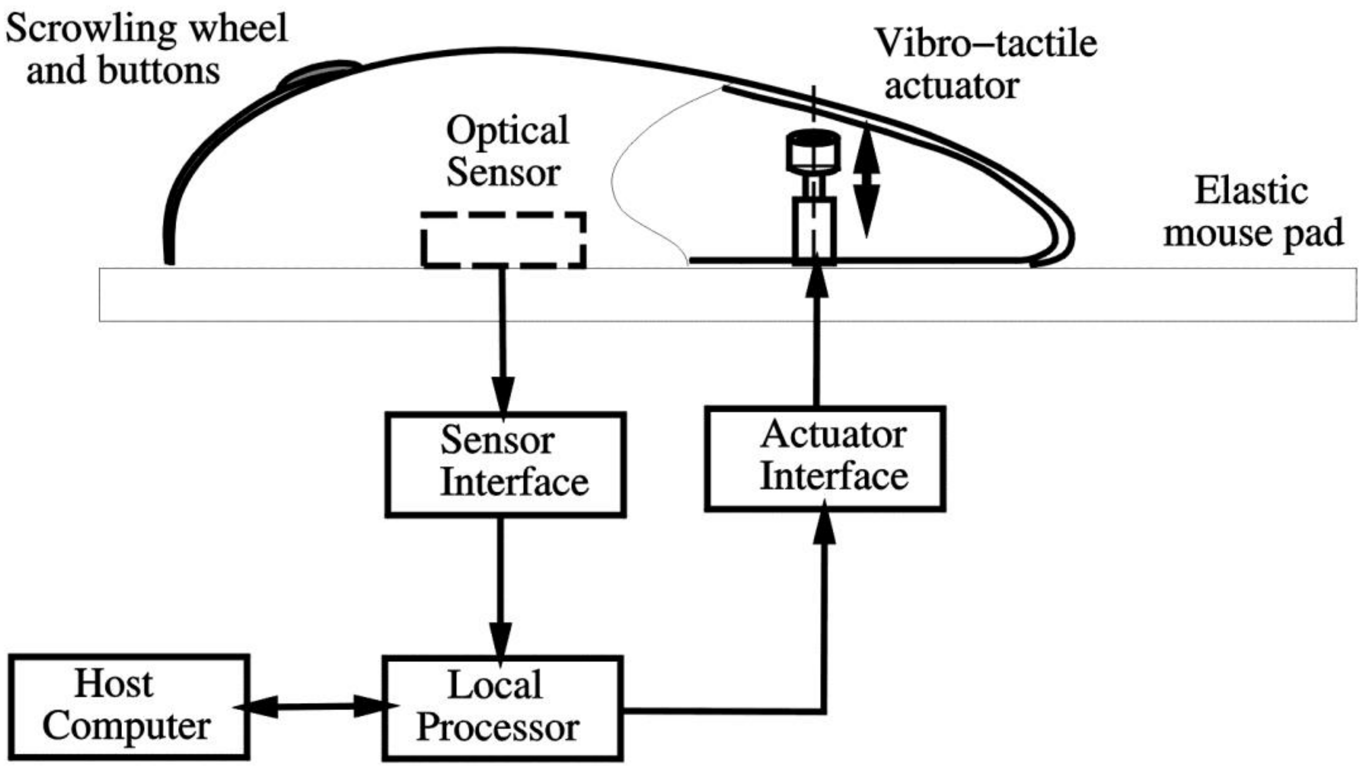
Temperature, vibration and skin deformation

Most widely used devices generate vibration (known as vibro-tactile)

Range from very simple such as mobile phones, to gloves with multiple embedded vibrating units

Cybernetics

Haptic Displays - Tactile



The iFeel Mouse (0-125 Hz).

Glove

CyberTouch Glove (Virtex): 6 individually controlled vibrotactile actuators
0-125 Hz frequency,
1.2 N amplitude at 125 Hz

Haptic Displays - Tactile

Devices that work by causing small deformations in the skin are usually based on arrays of mechanical pins which can be moved as required

Challenging – from earlier we know that we need an array with 1mm between pins and bandwidth >300Hz

Pins are usually actuated by electromagnetic methods: solenoids, voice coils etc. due to the high performance they are capable of (speed and force)

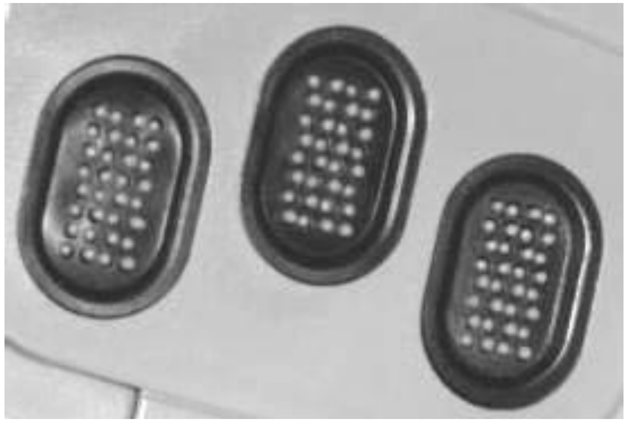
Very difficult to make portable. Generally big and heavy

Tactile Pin Array - Actuation

Principle	Description	Comments
Thermo-electric stim-ulators	Use shape memory alloys which rapidly change shape on change of temperature	Can exert high forces at large displacements but thermal capacity of material limits frequencies to 10Hz
Electro-Magnetic	Use solenoids or small motors	Such mechanical devices lose performance as their size drops resulting in small forces and low refresh rates.
Pneumatic	Use air pockets or plungers controlled by air valves	Can be made small, light weight and flexible (as air source elsewhere). Though compressibility of air limits frequencies to around 10Hz
Piezo-electric	Piezoelectric actuators move lever mechanisms	Can display high forces at a wide range of frequencies but lever mechanisms at required density result in large devices.

Haptic displays – Tactile pin arrays

Full display

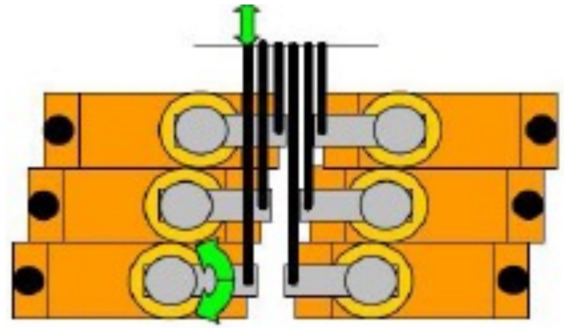


6*6 display showing sinewave

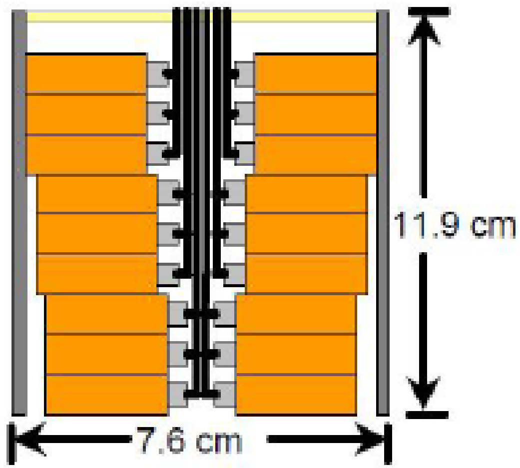
Cybernetics

Servo Arrangement

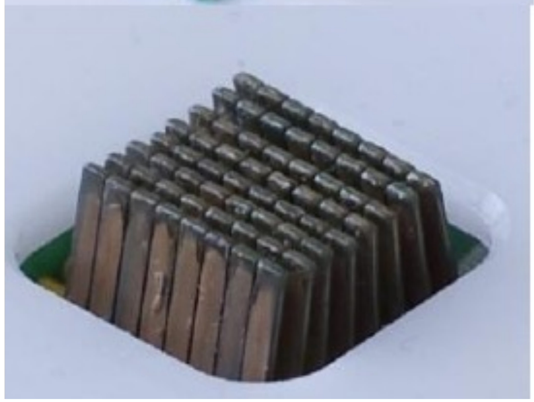
Servos tightly packed for
2mm pin spacing
Servo rotates translating
to vertical movement



Vertical layout of servos
Six blocks of six servos
rigidly attached to
aluminium chassis



Haptic Displays - Tactile



See laterotactile.com

Stimulator of Tactile Receptors by Skin Stretch (STReSS)

Lateral skin deformation (now laterotactile stimulation) was first described in (Hayward and Cruz-Hernandez, 2000).

Travelling wave of lateral skin deformation induces sensation of moving feature under finger tip.

STReSS is a 2D laterotacltile display … stimulates finger tip with matrix of piezoelectric actuators

Haptic Displays – Force Feedback

Devices which provide force back as well as touch

Used in research since the 60's and available commercially for 20 years.

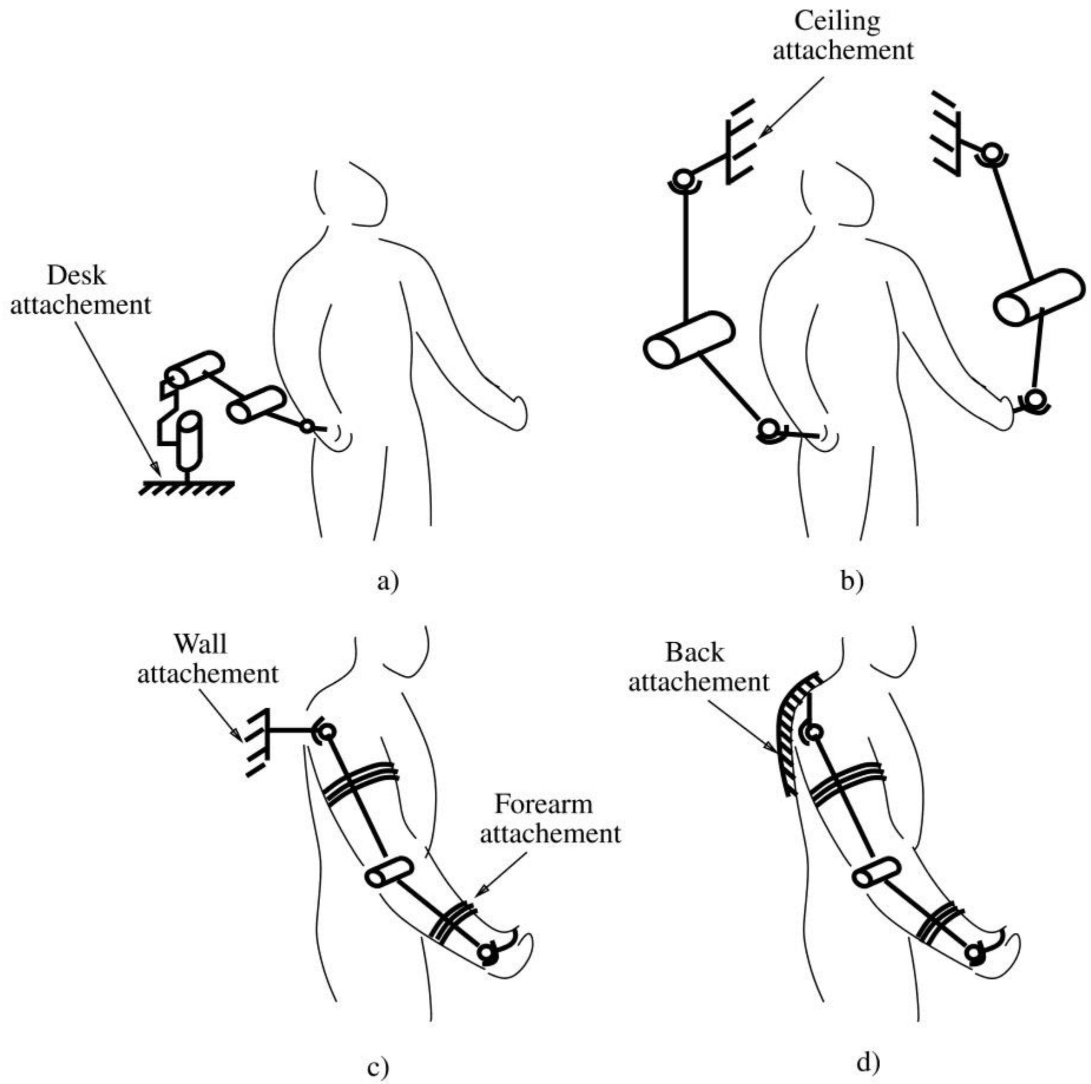
Until recently, prohibitively expensive and experimental

Haptic Displays – Force Feedback

Generally need mechanical grounding to resist user motion

- Exceptions: using gyroscopic forces or air jets
- But limited applications

Can be grounded on desk, wall, or on user body;



Haptic Displays – Force Feedback

Different ways to compare performance, some are:

- Apparent Mass: How much inertia the device has (not just weight)
- Workspace: How large the interactive area is
- Stiffness: How rigid a virtual surface can be displayed

Generally speaking can divide force displays into back-drivable and admittance controlled

FF - Back-drivable devices

Moved by the user

When in “free space” the motors are turned off.

When a virtual surface is contacted the motors turn on and resist movement

Designed to be as lightweight as possible

Trade-off is low force and low stiffness

FF : Admittance Controlled Devices

Use a force sensor to measure user input and move themselves out of the way.

When in "free space" the motors are running and responding as quickly as possible

When a virtual surface is contacted the motors turn off/resist movement

Designed to be as rigid and high force as possible

Trade-off is high apparent mass and low bandwidth.

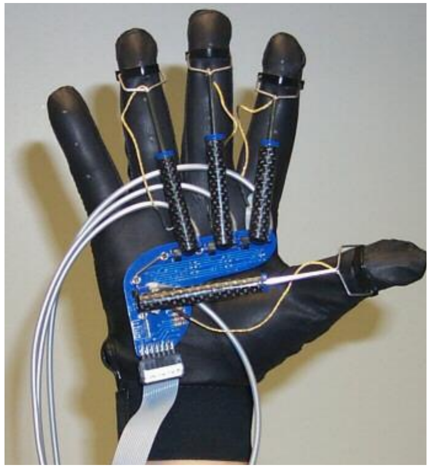
Force Feedback Examples - Gloves

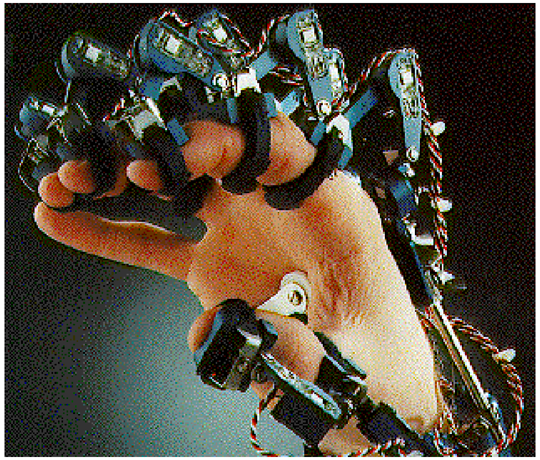
Pros:

- Many points of contact
- Lightweight so quick to move
- Very large workspace

Cons:

- Limited force direction per contact
- Slow force response
- Low stiffness
- Only basic geometry displayed, no weight or lateral forces (friction)





Serial Linkage Admittance Controlled

Pros:

- High forces
- Low mass and friction for smooth (low bandwidth) movements
- Large workspace
- High stiffness (can exceed human perception of stiffness)

Cons:

- Difficult to have multiple points of contact
- Can't make quick (high bandwidth) movements

Serial Linkage Back-drivable

Pros:

- Usually light/low mass
- Moved very quickly
- Simpler to build & control
- Can have multiple contacts each with multiple DoF

Cons:

- Small motors and light linkages mean low forces and low stiffness.
- Usually small workspaces

Parallel Linkage Back-drivable

Pros:

- More rigid structure than serial linkage, so higher achievable stiffness
- Also higher force output

Cons:

- Small workspaces
- Difficult to have multiple points of contact

Magnetic levitation

Pros:

- High force and high stiffness
- Very low friction

Cons:

- Very small workspace
- Difficult to construct and control

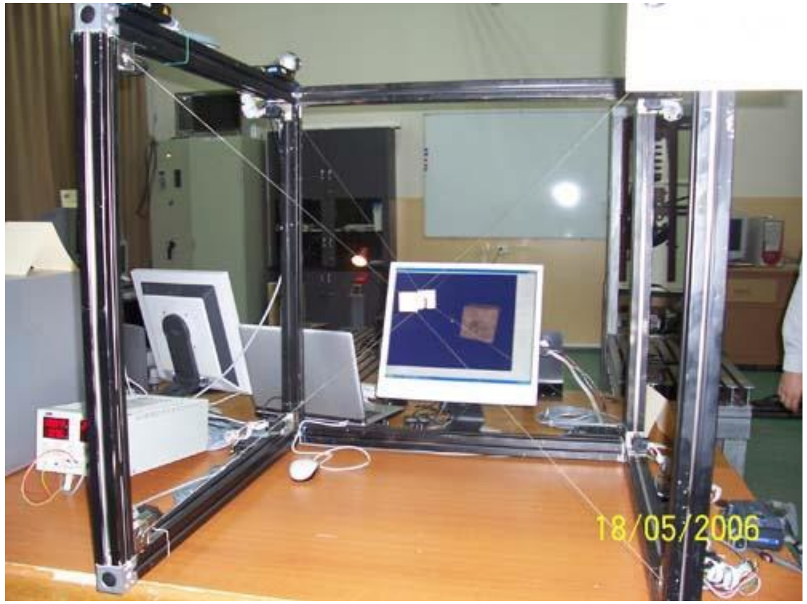
Parallel String/Cable drive (SPIDAR)

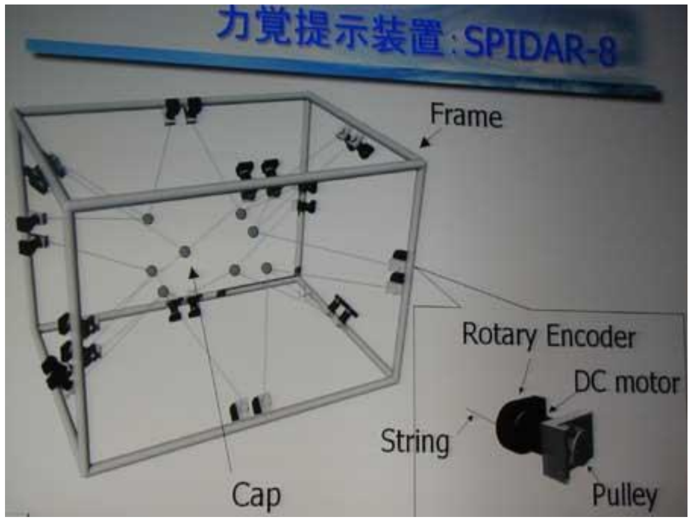
Pros:

- Low mass and good workspace
- Cheap/Simple

Cons:

- Low force output
- Can't move too quickly or string gets tangled
- High maintenance





Cybernetics

Degrees of Freedom

The Degrees of Freedom of a robotic device are: "The number of independent linear or rotational movements that completely specify the position and orientation of the whole mechanism"

In a haptic device:

- Active DoF can be controlled by the device itself (usually by motors)
- Passive DOF can be tracked, moved by the user but no forces felt

Degrees of Freedom

- The Human arm has ~ 7 DoF
- A free floating object in space has 6 DoF
- When controllable DOF of a robot = movable DOF it is known as holonomic
- A 6 DOF haptic interface is significantly more mechanically complex and hence expensive
- When are 6 DOF devices necessary?

Haptic Presentation Properties

Cues provided

Grounding

Number of display channels

Degrees of freedom

Form

Fidelity

Spatial and temporal resolution

Latency tolerance

Size/available workspace

Logistic Properties

user mobility

interface with tracking methods

environment requirements

associability with other sense displays

portability

throughput

encumbrance

safety

cost

Summary Questions

1. A touch screen interface is not usually considered haptic. Why not and what might be added to make it into a haptic device?
2. It is required to design a full body tactile feedback suit. What aspects of the tactile sensory system can be exploited to make the design simpler?
3. A training simulator is to be built to teach astronauts to assemble equipment in zero gravity. A body grounded force feedback glove system is being considered, what are the pros and cons of using this type of device.