Multimedia Systems

About the Lecturer

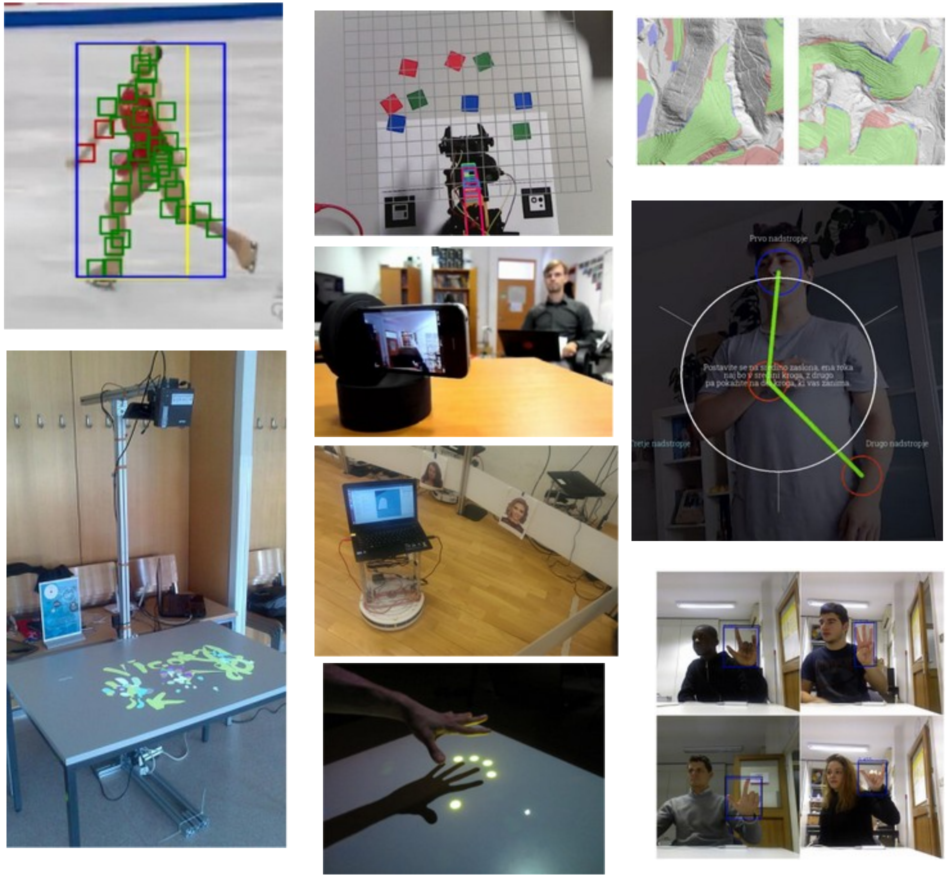
Luka Čehovin Zajc, PhD

Assistant Professor

Visual Cognitive Systems Laboratory

Office R2.39

luka.cehovin@fri.uni-lj.si



Course requirements

- **Laboratory exercises / project work - 50%**
  - Practical exercises - grading throughout semester
  - Single project – grading at the end of the semester
  - *Only valid for the current school year*
- **Exam (written + oral) -50%**
  - Must pass laboratory exercises to attend
  - Theoretical and practical assignments
  - Optional oral exam for borderline students (50% to ~65%)
  - Only oral exam for less than ~10 students

Laboratory exercises

- Teaching assistant: **Me**
- Practical consolidation of selected topics
  - Python (Jupyter, SciKit, NumPy, ...)
  - Hosted Jupyter instances at lab.vicos.si
  - Local installation (virtualenv, Docker)
  - Google Colab
- Each exercise is due in two weeks (approximately)
  - Timely assignment hand-in encouraged
  - Labs = Presentation + Consultations + Defenses

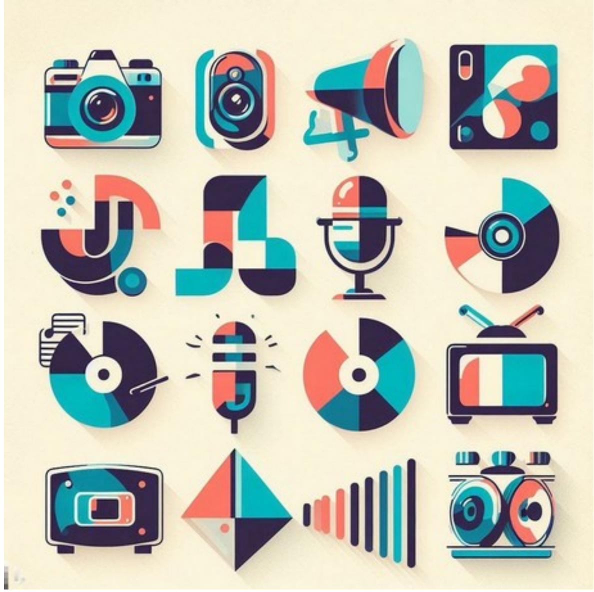
Project assignment

- Alternative to regular laboratory exercises
- In-depth project work on a selected topic
  - You have to pace your work yourself
  - Meetings can be arranged to discuss topic
- Work has to be finished by the end of semester
  - Presentation in classroom
  - Demonstration
  - Code hand-in

Project topics

- Sketch-based Image Retrieval
- Image blending using deep learning
- Deep learning for compression

Contact me
if you are interested!



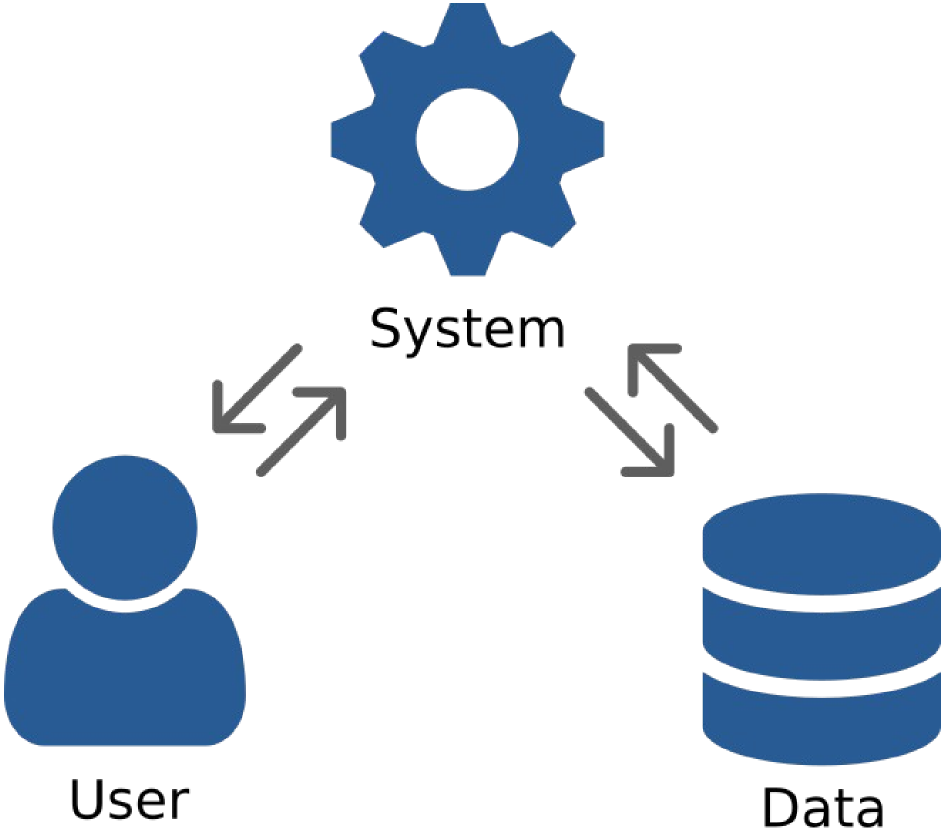
multimedia *(Latin)*

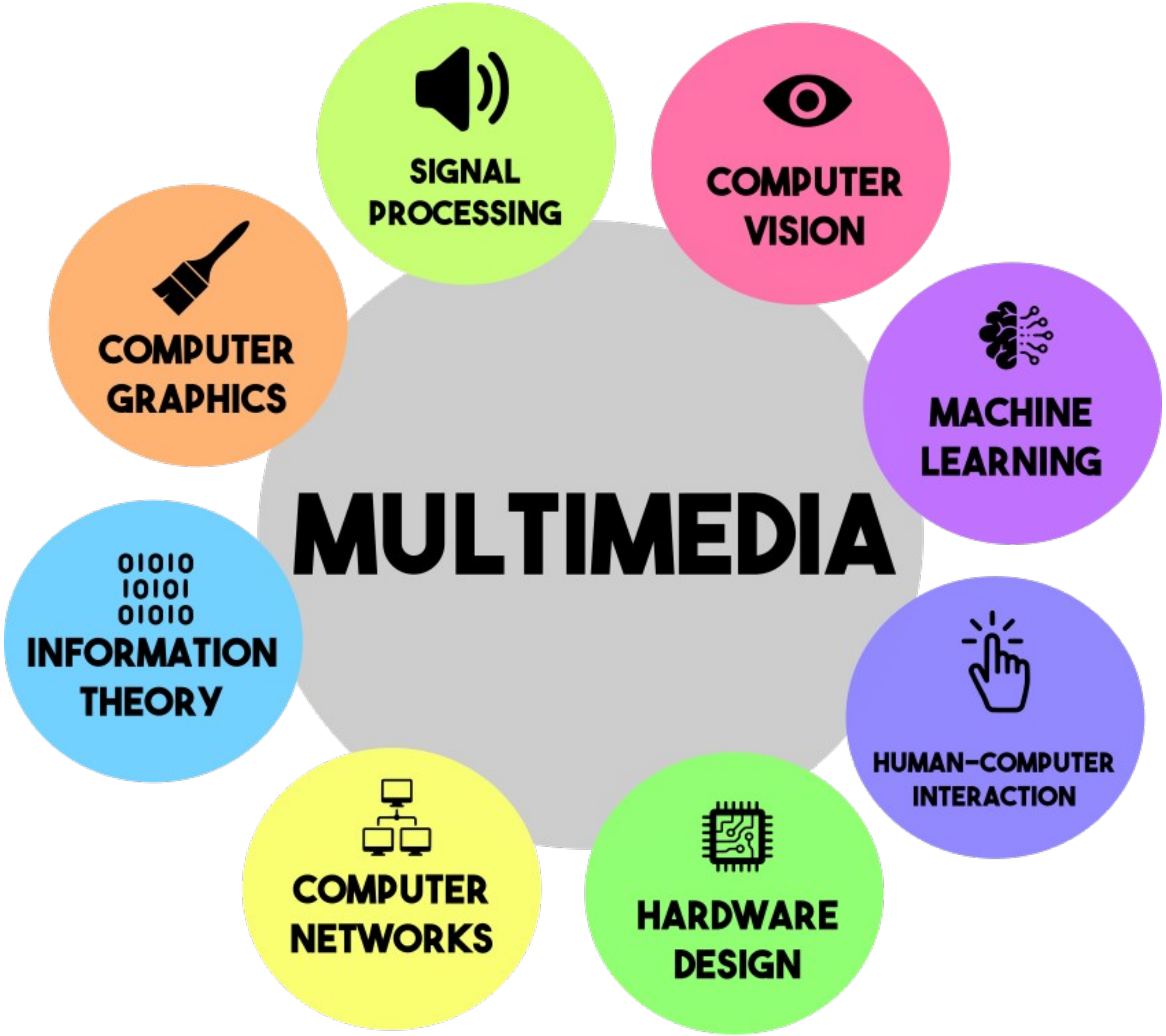
multum + medium

Hypermedia

- Ted Nelson (~1965): HyperText
  - Book: linear medium
  - HyperText: non-linear (interactive)
- Hypermedia: not only text
  - Form of multimedia application
  - WWW – type of hypermedia application

Multimedia Systems





MULTIMEDIA

SIGNAL PROCESSING

COMPUTER VISION

MACHINE LEARNING

HUMAN-COMPUTER INTERACTION

HARDWARE DESIGN

COMPUTER NETWORKS

INFORMATION THEORY

COMPUTER GRAPHICS

Application domains

- Digital television, video on demand (video + sound)
- Computer games (graphics + sound + interactivity)
- Teleconferences (video + sound)
- Remote lectures (video + sound + slides)
- Telemedicine (video + sound + haptic + manipulation)
- Large databases (e.g. Google, YouTube, Facebook, Amazon, Dropbox)
- Extended reality
- Data visualization (image + sound + interactivity)

Research challenges

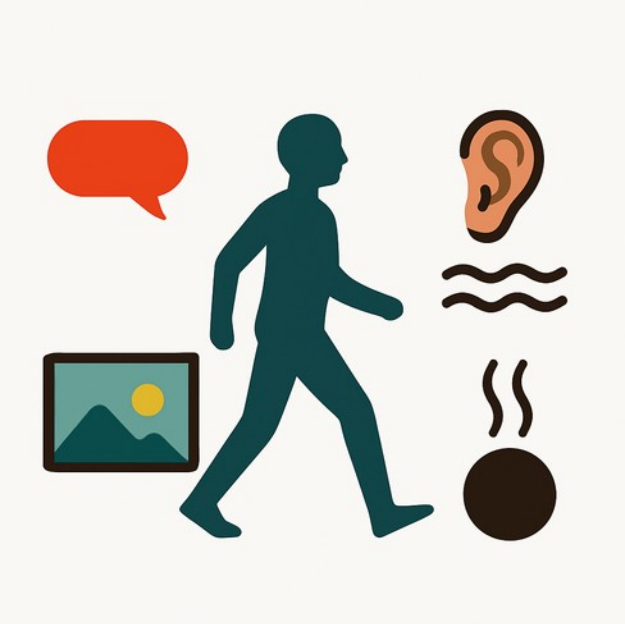
- **Processing**
  Content analysis, information retrieval, enhancement, etc.
- **Storage, transmission**
  Quality of service, compression, security, IO devices, etc.
- **Tools, applications, methods**
  Content manipulation, user interfaces, multi-modal interaction, content production systems, collaboration systems, etc.

Lectures overview

- Human Perception
- Signal Processing
- Multimedia Compression
- Information Retrieval
- Storage and Networking
- Hardware and Emerging Technologies

Supporting Literature

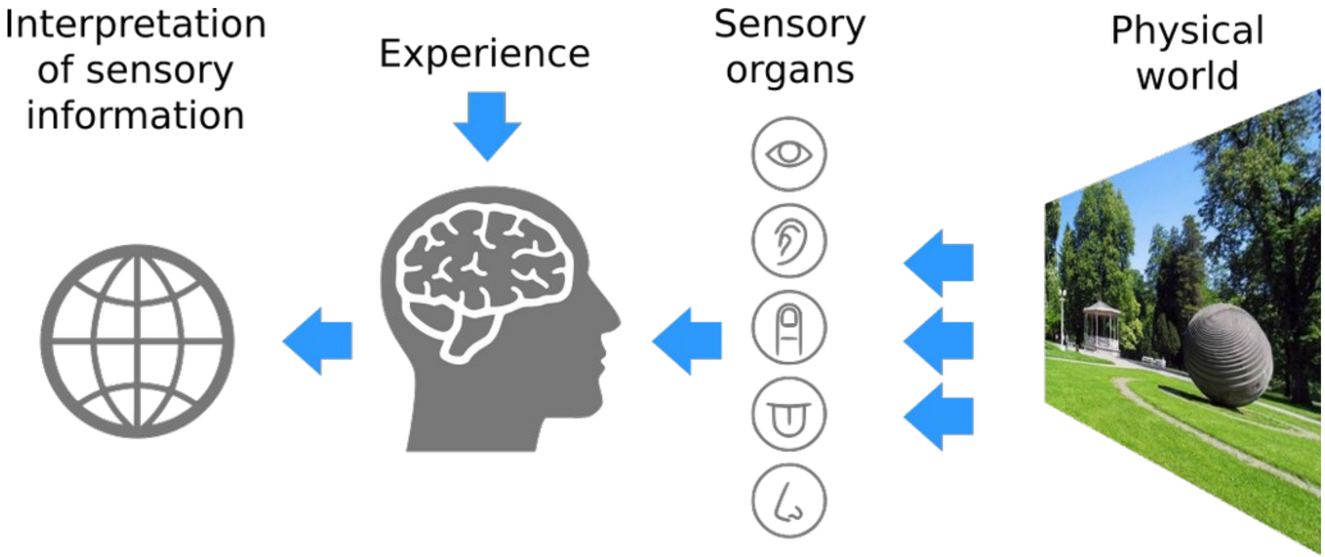
- Slides + lecture notes available at online Classroom (Učilnica)
- Li Ze-Nian, M. S. Drew, Fundamentals of Multimedia, 2010 - *overview, general topics*
- C. D. Manning, P. Raghavan, H. Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. - *information retrieval concepts*
- Gonzalez and Woods: Digital Image Processing
- J. O. Smith III, Introduction to Digital Filters



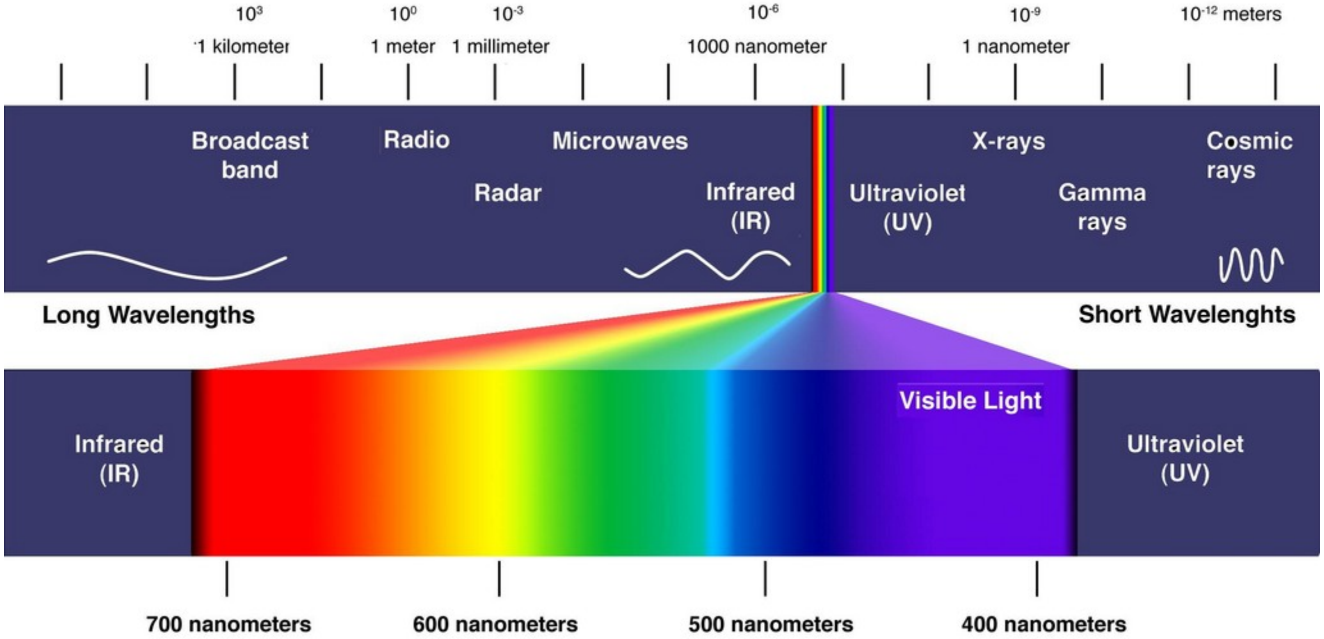
Humans and Multimedia

Human Perception

- Reality is subjective
  - Sight
  - Hearing
  - Haptics
  - Chemical
  - Balance



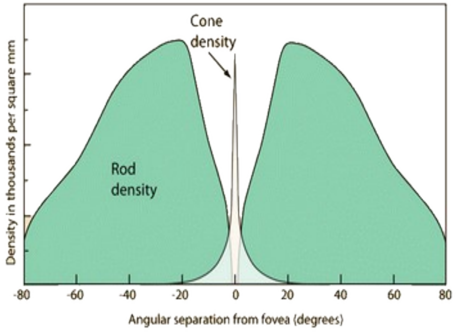
About Light

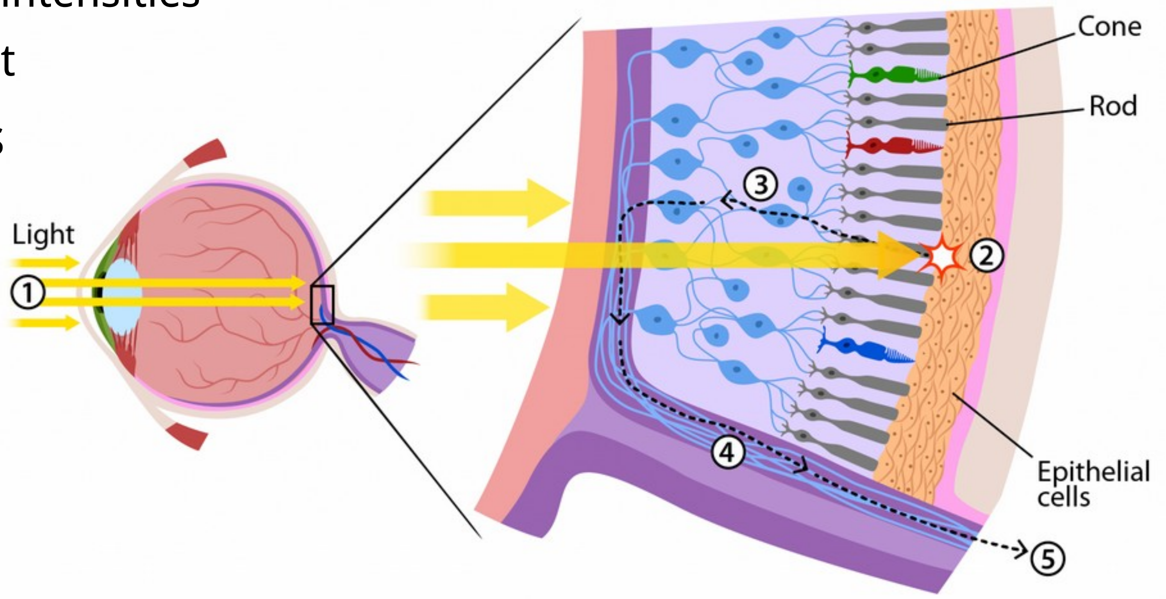


- Light is electromagnetic waves / particles (photons)
- Visible light is light with wavelength from ~400nm to ~700nm

Perceiving Light

- Eye perceives light that falls on the retina
- Retina is composed of two types of cells
  - Cones - Sensitive to color and large intensities
  - Rods - Sensitive to low intensity light
- There are more rods than cones
- Not uniform distribution





Why are we Trichromatic?

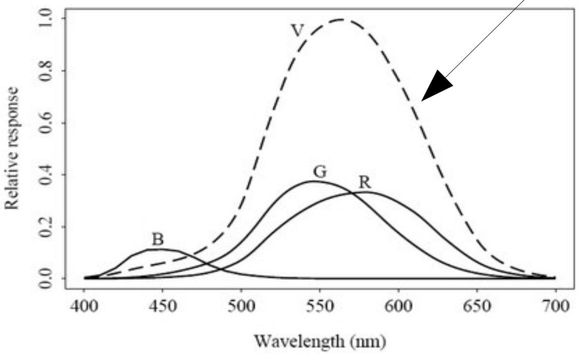
- Young-Helmholtz theory (19th century)
- Three types/lengths of cones
  - Different wavelengths (R=L, G=M, B=H)
- It is not yet entirely clear how brain combines color information
  - Ganglion trigger to differences R-G, G-B, B-R (opponent theory)
- All three channels are combined into achromatic information

Spectral Sensitivity of the Eye

- Eye is most sensitive to the middle of visible spectrum
- Cone distribution approximately R:G:B == 40:20:1 (varies from human to human)
- Rods are more sensitive to wavelengths closer to the red part of the spectrum.

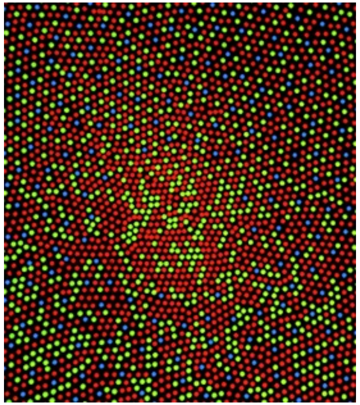
Cones sensitivity

The curve for blue is not plotted on the correct scale, it is much lower than the curve for red or green.



Rods sensitivity

Sensitivity of rods is similar to the overall sensitivity curve V for cones, it is only shifted towards the red spectrum.

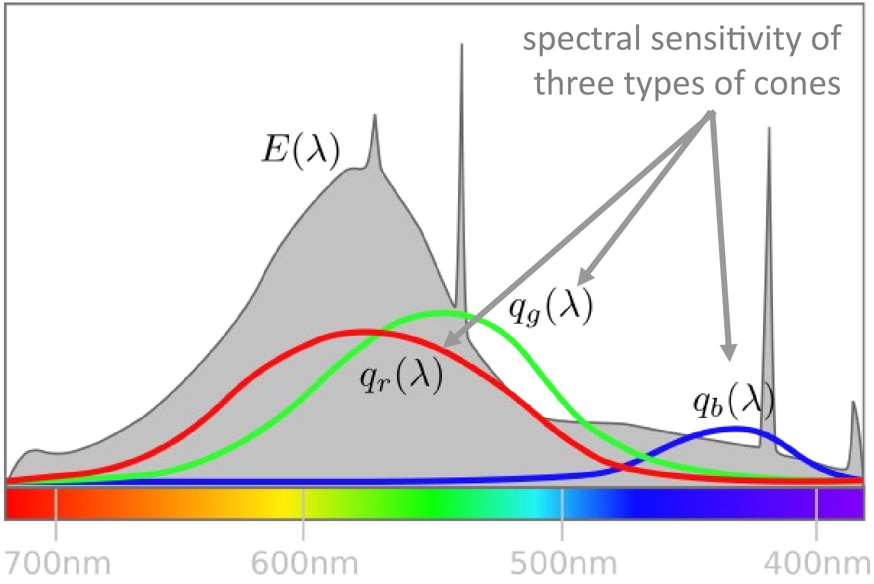


Cone distribution

Cones Sensitivity

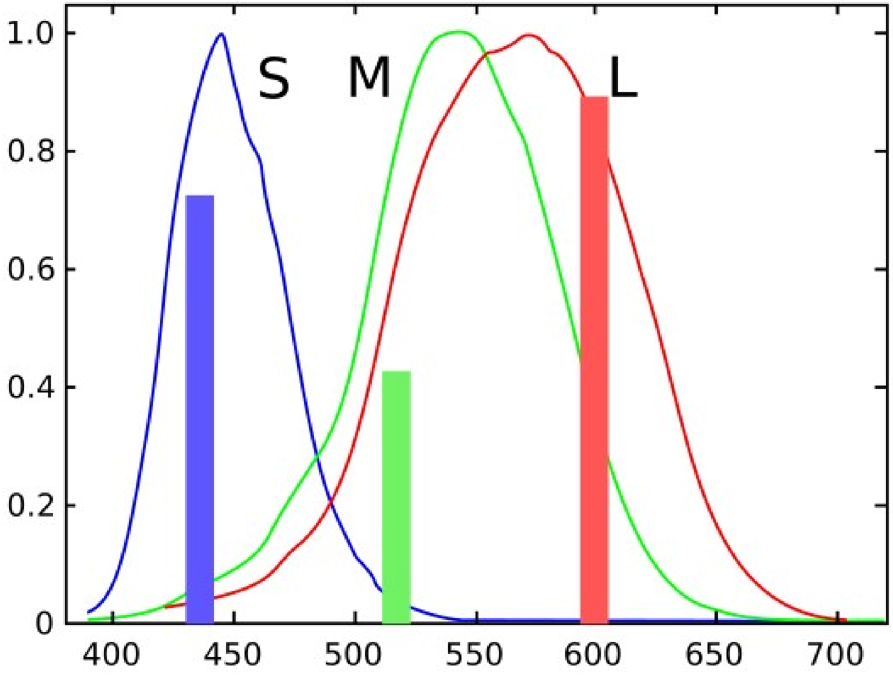
- Cones are triggered with different intensity with respect to the light's wavelength
- Filtering color spectrum $E(\lambda)$

$$R = \int E(\lambda) q_r(\lambda) d\lambda$$
$$G = \int E(\lambda) q_g(\lambda) d\lambda$$
$$B = \int E(\lambda) q_b(\lambda) d\lambda$$



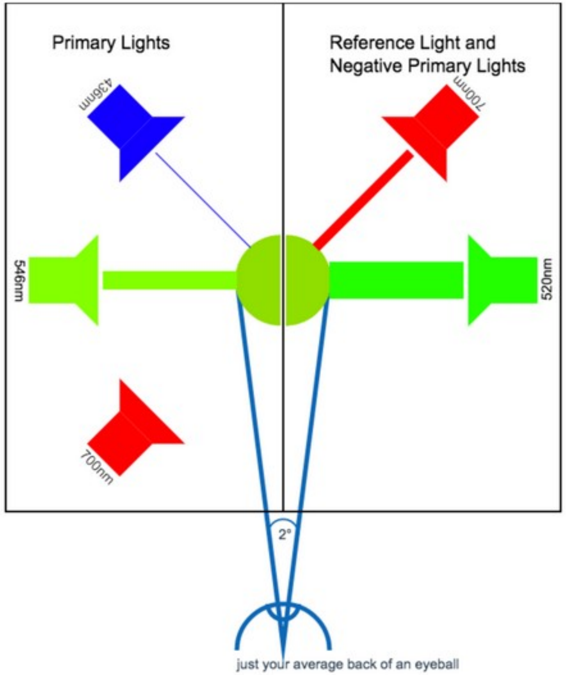
Simulating Color

- Stimulating cone cells
- Metamerism
- Color primaries
  - Trichromatic (3+)
  - Different standards

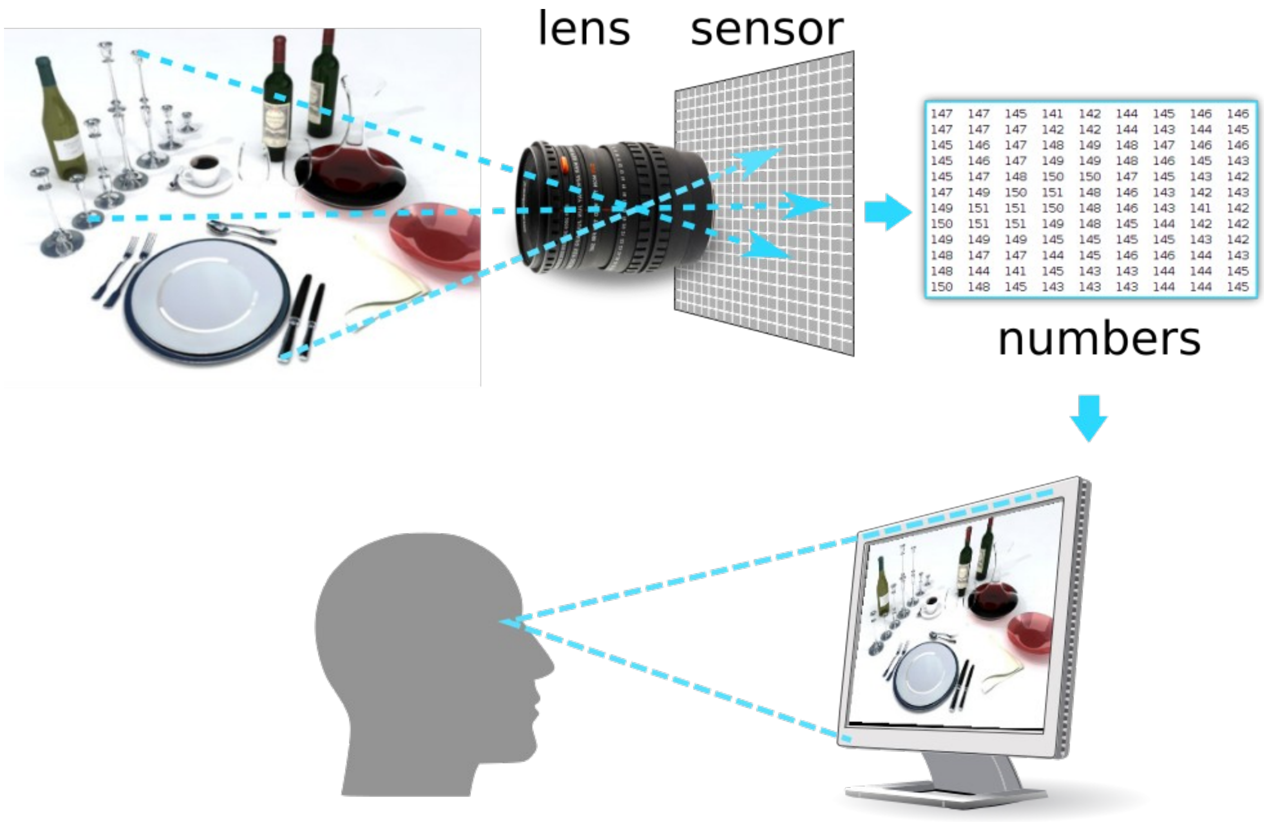


Measuring color perception

- Color reproduction evaluation
- Quantitative evaluation it in terms of human perception
- The tristimulus colorimeter experiment
  - Matching reference color
  - A person is controlling the intensity of three color channels
  - Standard observer (field-of-view)
  - Negative light


Primary Lights
Reference Light and Negative Primary Lights
436nm
700nm
546nm
520nm
700nm
2°
just your average back of an eyeball

World, image, eye



147	147	145	141	142	144	145	146	146
147	147	147	142	142	144	143	144	145
145	146	147	148	149	148	147	146	146
145	146	147	149	149	148	146	145	143
145	147	148	150	150	147	145	143	142
147	149	150	151	148	146	143	142	143
149	151	151	150	148	146	143	141	142
150	151	151	149	148	145	144	142	142
149	149	149	145	145	145	145	143	142
148	147	147	144	145	146	146	144	143
148	144	141	145	143	143	144	144	145
150	148	145	143	143	143	144	144	145

About Sound

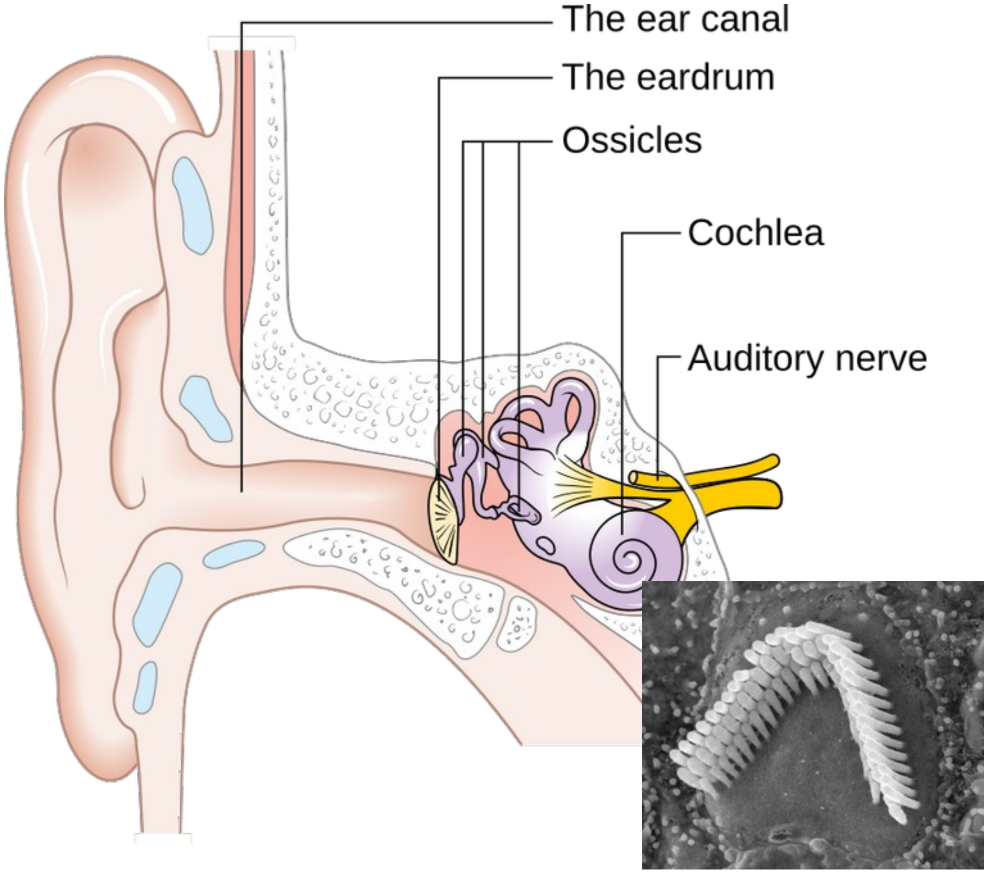
- Wave of pressure in medium
  - Particles repeatedly compressed and expanded
  - Longitudinal waves
  - Requires medium (air, water)
  - Electronic representation - audio
- Wave phenomenon
  - Reflection – bouncing
  - Refraction – angle change when entering different medium
  - Diffraction – bending around obstacle

Measurable Sound Characteristics

- Frequency (Hz)
  - Number of occurrences of a repeating event per unit of time
- Amplitude, pressure, intensity (W/m^2)
  - Amount of change over a period
- Duration (seconds)
- Direction
- Speed
  - Speed based on medium
  - Air: ~331 m/s

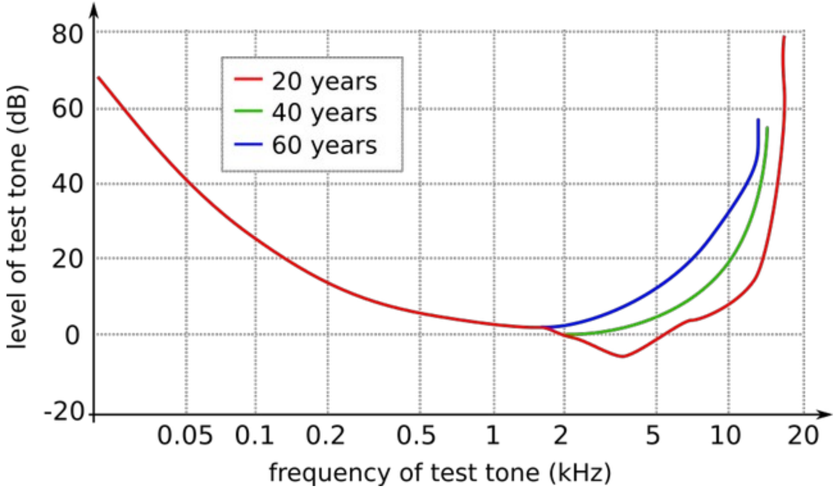
Human auditory perception

- Sound travels the ear canal to the eardrum that vibrates
- Ossicles amplify the vibration
- Cochlea contains liquid that vibrates
- Liquid shakes hair cells
- Hair cells are sensitive to different frequencies
- Responses are transmitted via auditory nerve



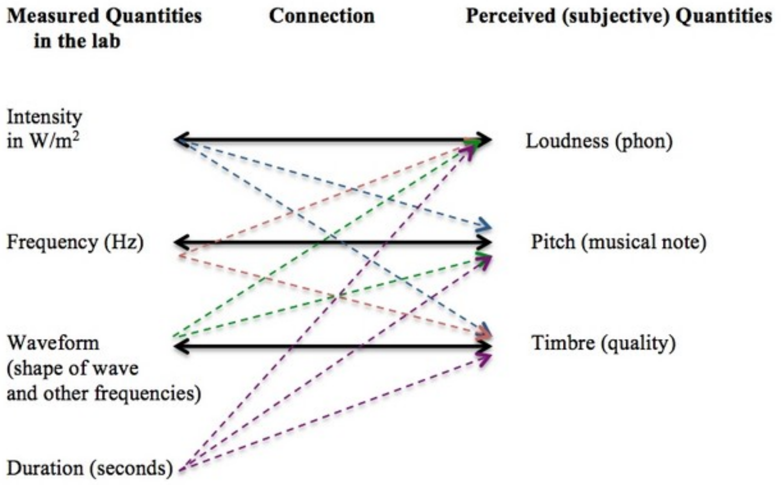
Human ear sensitivity

- Frequencies between 20Hz and 20kHz
  - Some have to be louder than other
- Threshold of hearing
  - Amplitude where a pure tone is detected with 50% accuracy



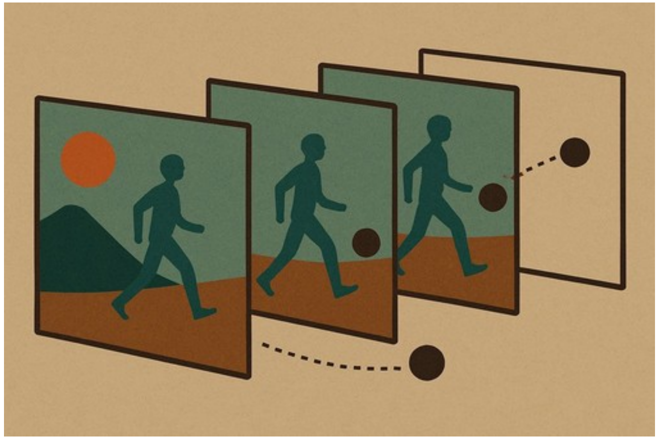
Perception of sound

- Pitch (low/high)
- Loudness (loud/soft)
- Timbre, tone color
  - Combination of multiple frequencies
  - Change over time
- Sonic texture
  - Multiple sources
  - Unison, polyphony, homophony, cacophony
- Spatial location



Sensing in Space and Time

- Sight
  - Changing scene
  - Temporal resolution
  - Integration of motion
- Hearing
  - Spatial sound
  - Echolocation



Video Temporal Resolution

- Human perception system (eye+brain) can perceive about 10 - 12 images per second as separate images.
- Persistence of vision
  - Image „remains“ in cortex for 1/25s
  - Neuron saturation

Sight and Hearing Characteristics

	Sight	Hearing
Spatial	Fovea density: ~150,000 cones/mm^2 1 m distance = about 0.3 mm detail	Horizontal (azimuth): ~1–2° Vertical (elevation): ~10°–15° Poor distance perception
Temporal	Photopic (well-lit) conditions 50–60 Hz	Sensitive to fine temporal structure Range: 20 Hz – 20 kHz Discrimination ~2 Hz at 1 kHz

Touch

- Pressure, vibration, texture, and temperature
  - Mobile and Wearable computing (e.g., touchscreens, VR controllers, haptic feedback in phones).
  - Research challenges: realism of tactile rendering, latency, and multi-point contact.

Taste and Smell

- Chemical senses
- Rarer in multimedia, but experimental systems exist (digital scent displays, olfactory VR).
- Smell is strongly linked with memory and emotion, which makes it powerful for immersion.
- Currently confined to research labs or niche entertainment.

Balance and Proprioception

- Balance
  - Inner ear detects acceleration, rotation, and spatial orientation
  - Mismatches between visual input and vestibular signals cause motion sickness in AR/VR
- Proprioception
  - Awareness of limb position and movement, even without vision
  - Important in VR and motion capture; haptic suits and exoskeletons attempt to stimulate it.

Representations and Multimedia

Representations

- A structured way of describing real-world information so it can be stored, processed, or transmitted
- Emphasizes some aspects, ignores others
- Describe a city
  - A map
  - A photograph
  - GPS coordinates

Mathematical Formulation

- Vector of values
  - Encoding data properties
  - Embedding – special case (structured space)
- Task dependent
  - Select the right representation
  - General vs. specialized
  - What to describe?

Levels of Abstraction

Type	Image	Audio	Video
Low	Pixel	Waveform	Pixels
	Edges	Spectrogram	Flow
	Shapes	Phonemes	Correspondences
	Object	Words	Objects
High	Scene	Meaning	Actions

Examples of Representations

- Image
  - Histogram, average color
  - Objects, relationships
- Video
  - Sequence of images
  - Trajectories
  - Semantic Actions
- Audio
  - Waveform
  - Spectrogram
  - Generator parameters
- Text
  - Word frequency
  - Intention

Representations in Machine Learning

- Learning suitable representations automatically
- Deep learning
  - Layers of increasingly abstract features / representations
  - Embedding / latent space
- Use cases
  - Transformations, generation
  - Retrieval
  - Summarization, understanding

Summary

- Multimedia exists to communicate with our senses
  - Perceive the world – encode with representations
  - Decode representations – stimulate senses
- Understanding perception + representation helps us design systems that are:
  - More efficient
  - More natural to use
  - More immersive