Unidade Curricular

“Informação e Codificação”

António José Ribeiro Neves

an@ua.pt

https://www.ua.pt/pt/uc/15264

universidade de aveiro

ECIU university

deti

IEETA

LASI

Outline

Lossy compression

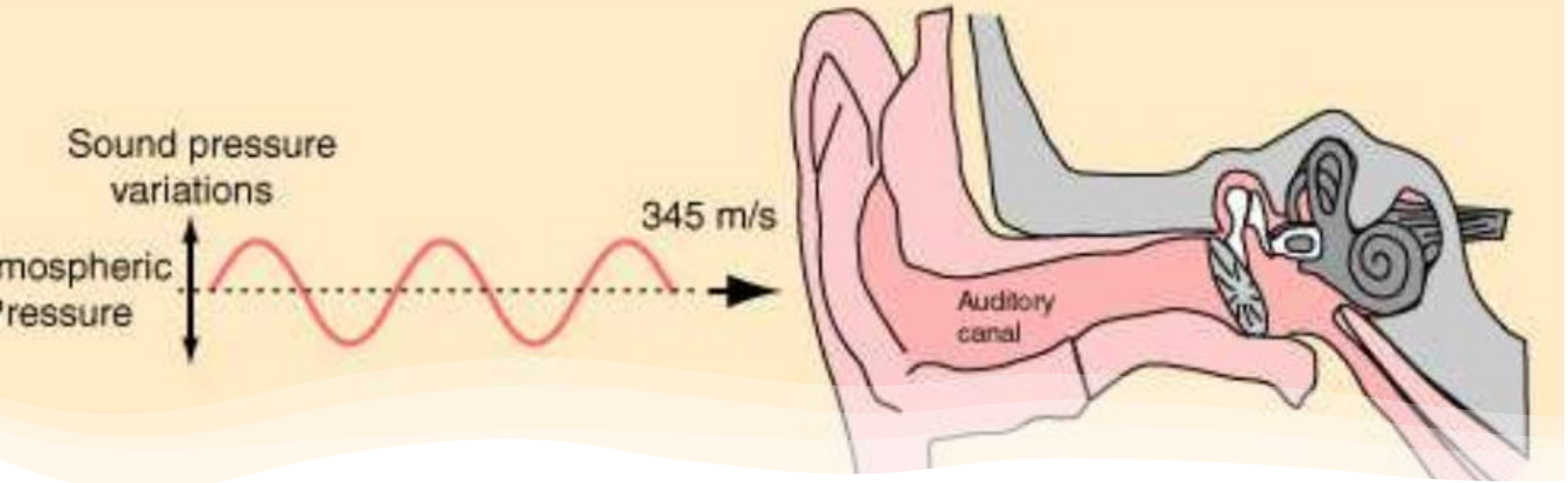
Guiding questions

Human Visual System (HVS)

Psychoacoustic model

Guiding questions

- What is the main difference between lossy and lossless compression?
- Why is understanding human perception important for lossy compression?
- What is the psychoacoustic model, and why is it critical in audio compression? (explore the topics of frequency and temporal masking).
- How does the Human Visual System (HVS) influence image compression?
- How do psychoacoustic and HVS models are used in the standards?

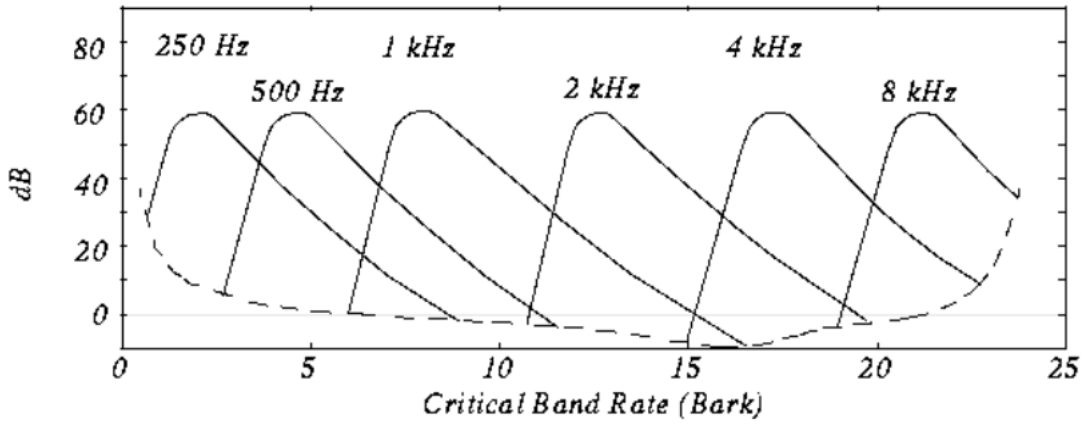


The human auditory system

- Humans perceive sound by the sense of hearing. By sound, we commonly mean the vibrations that travel through air and are audible to humans.
- Audio is the electrical representation of sound.
- Generally, humans can perceive variations in sound pressure from 16-20 Hz to 20-22 kHz.

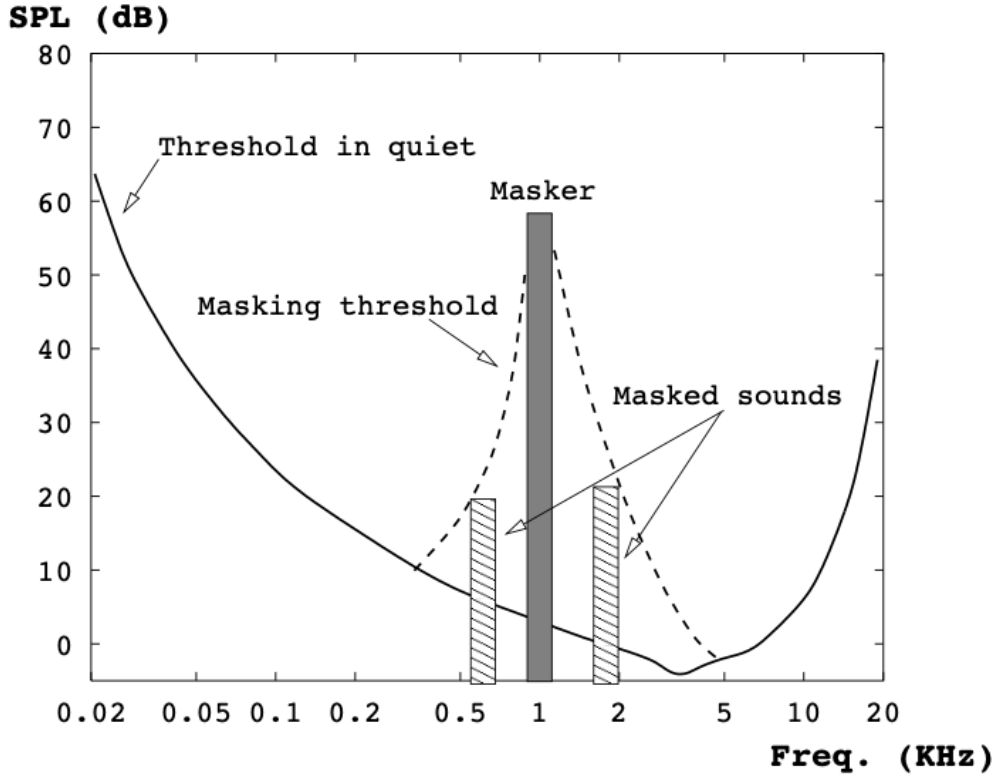
The human auditory system

- The auditory system can roughly be described as a bandpass filter-bank, consisting of strongly overlapping bandpass filters.
- These “filters” have bandwidths in the order of 50 to 100 Hz for signals below 500 Hz and up to 5000 Hz for signals at high frequencies.
- They are called critical bands.
- Twenty-five critical bands, covering frequencies of up to 20 kHz, are normally taken into account.



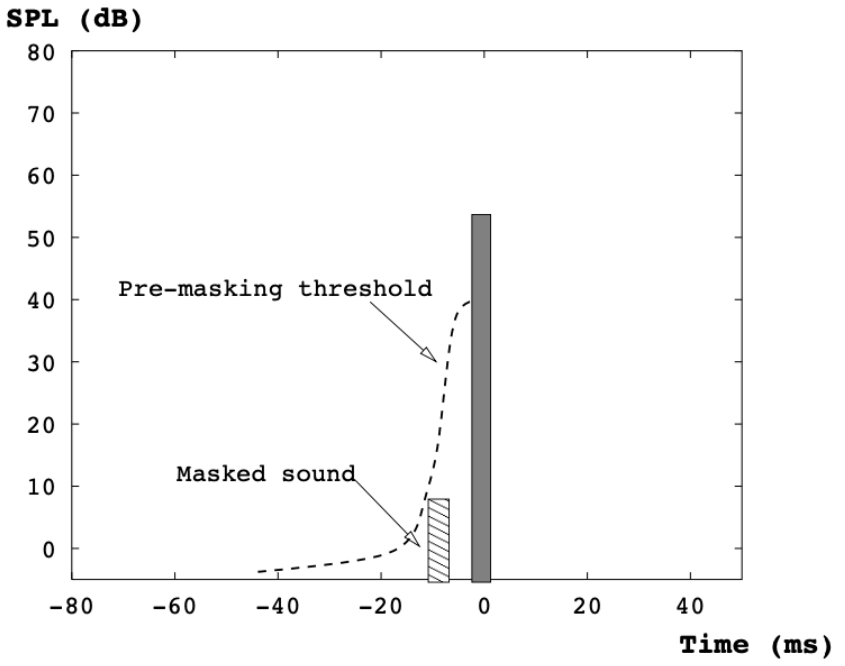
The human auditory system

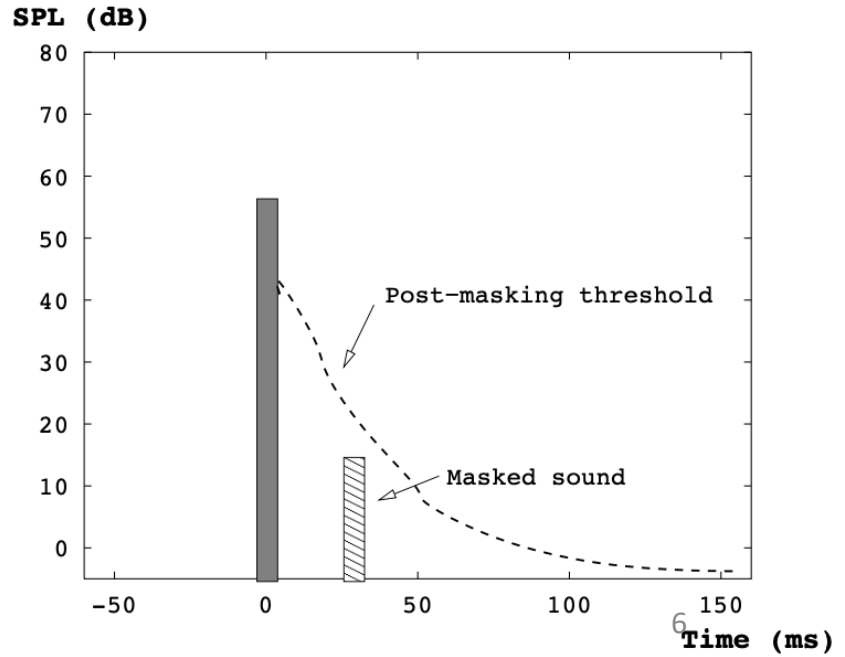
- Simultaneous masking is a frequency domain phenomenon where a low-level signal (maskee) can be made inaudible (masked) by a simultaneously and close in frequency stronger signal (masker).
- Such masking is greatest in the critical band in which the masker is located, and it is effective to a lesser degree in neighboring bands.



The human auditory system

- In addition to simultaneous masking, the time domain phenomenon of temporal masking plays an important role in human auditory perception.
- Temporal masking may occur when two sounds appear within a small interval of time.
- Depending on the individual sound pressure levels, the stronger sound may mask the weaker one, even if the maskee precedes the masker. . .
- The pre-masking has a duration of about 5 to 20 ms.
- The post-masking has a duration of about 50 to 200 ms.





Quality assessment of audio

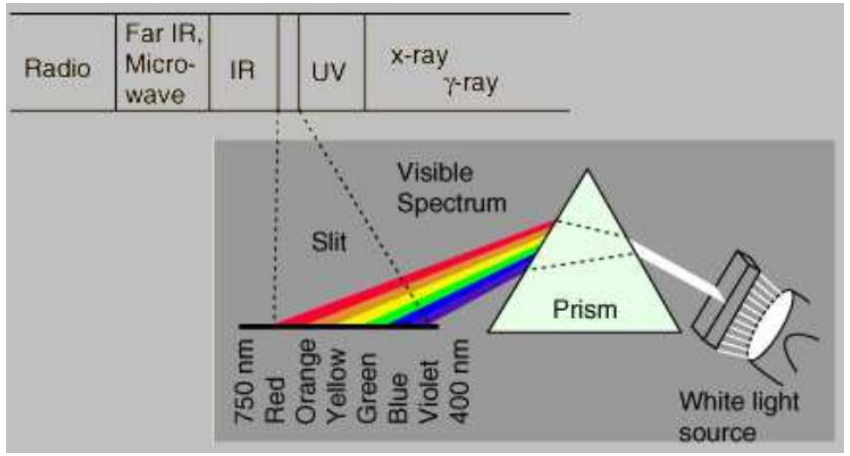
- The audio quality may be evaluated using subjective or objective measures.
- One of the scales used for subjective evaluation of wide band audio codecs is the ITU-R five grade impairment scale
  - 5.0 Imperceptible
  - 4.0 Perceptible, but not annoying
  - 3.0 Slightly annoying
  - 2.0 Annoying
  - 1.0 Very annoying
- Regarding objective measures, the signal-to-noise-ratio (SNR) is the most used.

The visible spectrum

Spectral colors (pure colors)

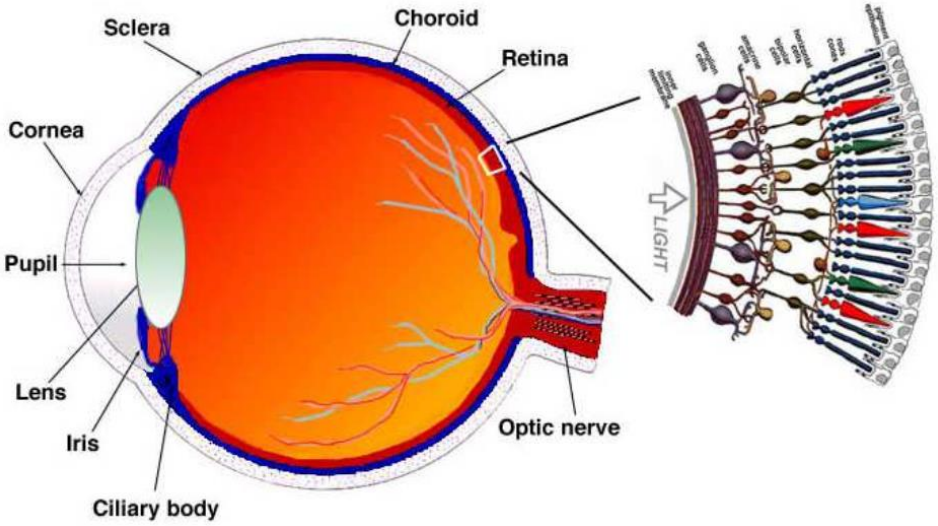
Cor	Wavelength
Violet	≈ 380–440 nm
Blue	≈ 440–485 nm
Cyan	≈ 485–500 nm
Green	≈ 500–565 nm
Yellow	≈ 565–590 nm
Orange	≈ 590–625 nm
Red	≈ 625–740 nm

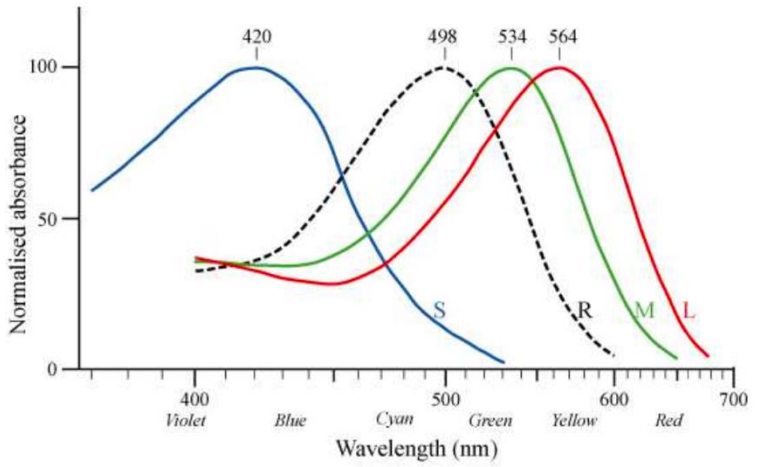
- The typical human eye senses electromagnetic wavelengths between 400 and 700 nm, and has maximum sensitivity around the 555 nm (green zone).
- Normally, the characteristics that allow colors to be distinguished are:
  - The brightness (how bright is the color).
  - The hue (the dominant color).
  - The saturation (how pure is the color).
- Together, the hue and the saturation define the chromaticity.
- Therefore, a color can be characterized by the brightness and the chromaticity.



The human perception of color

- The human vision is a complex process, still not completely understood, even after hundreds of years of research.
- The visualization of a physical process involves an almost simultaneous interaction of the eyes and the brain.
- This interaction is performed by a network of neurons, receptors and other specialized cells.
- The human eye is equipped with a variety of optical elements, including the cornea, iris, pupil, a variable lens and the retina.
- The human eye has photoreceptors that are sensitive to short wavelengths (S), medium wavelengths (M) and long wavelengths (L), also known as the blue, green and red photoreceptors.





R component

G component

B component

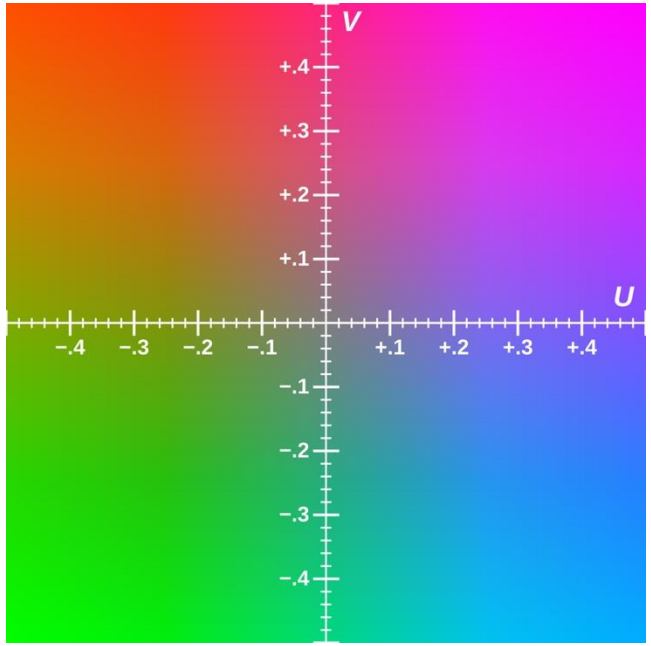
Color spaces (RGB)

C component

M component

Y component

Color spaces (CMY)



Y component

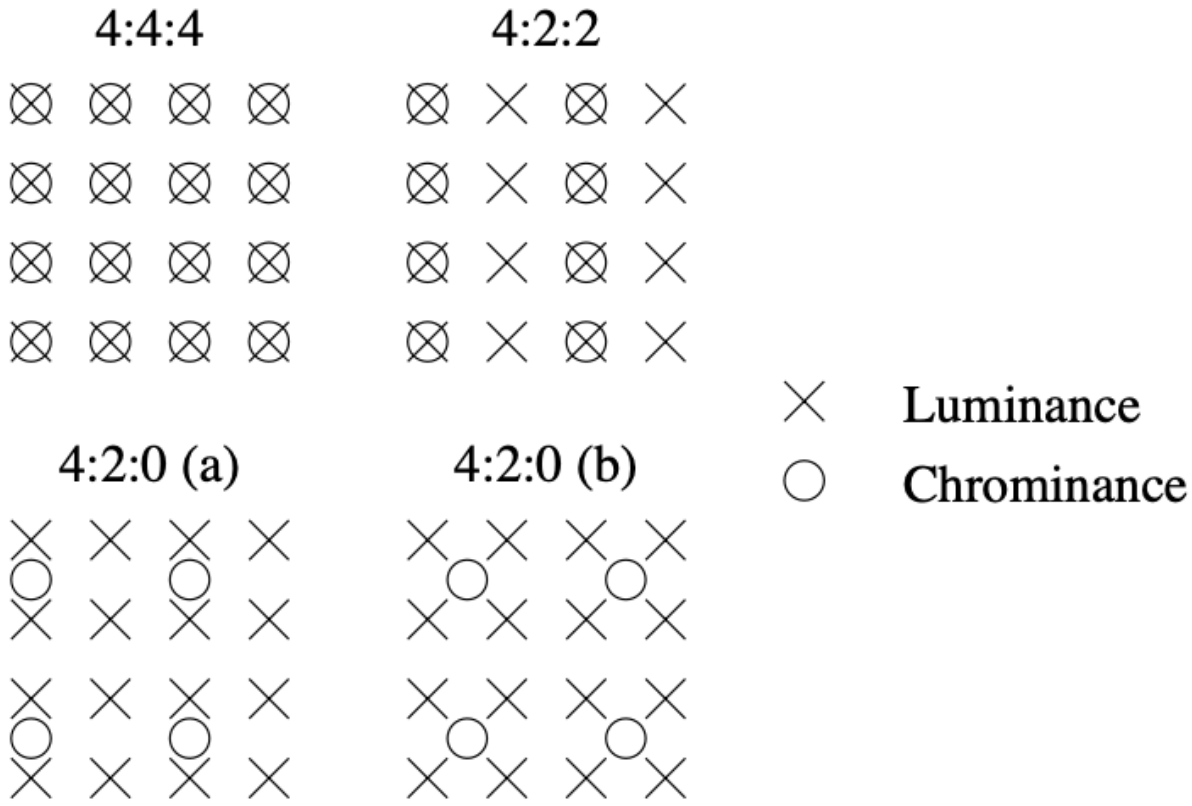
C_b component

C_r component

Color spaces (YUV)

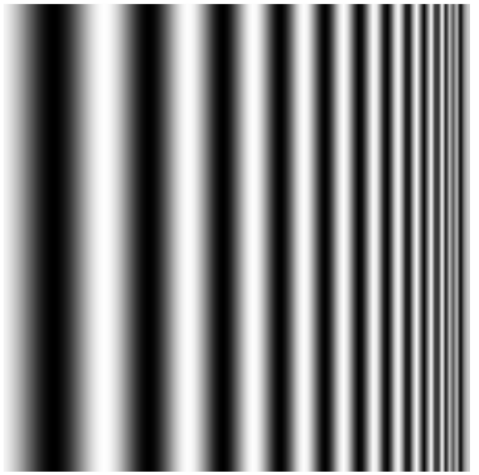
YUV color space

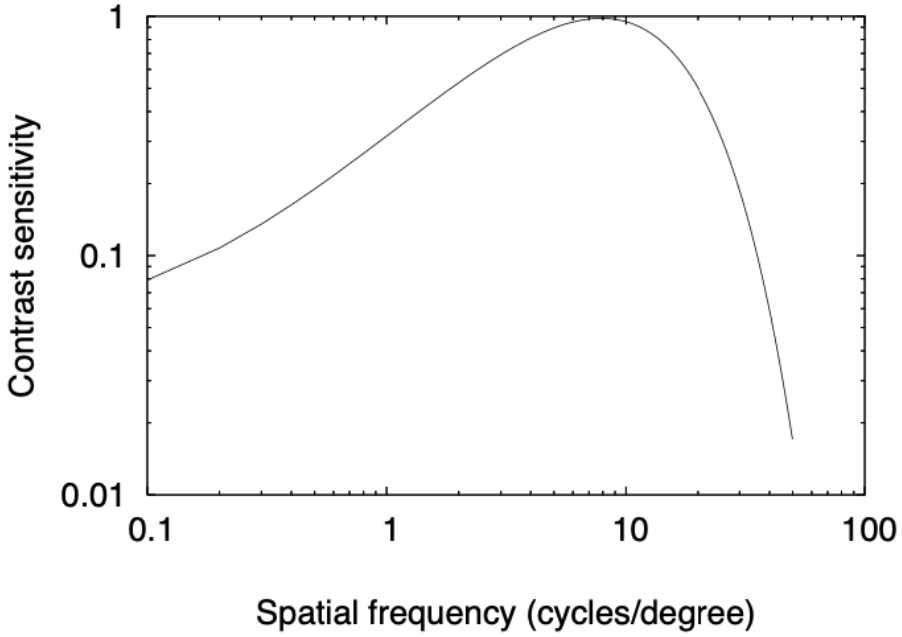
- The human eye is more sensitive in the green zone, which is represented mainly by the Y component (the U and V components are related to the blue and red.
- The YUV or YCbCr color spaces separate the chrominance component (UV / CbCr) from the luminance component (Y).
- Because the human eye is less sensitive to the blue and red, it is possible to reduce the bandwidth used to represent the U and V components, without introducing significant perceptual degradation.
- For this reason, it is common to sub-sample the chrominance components UV / CbCr , producing a reduction in the data rate.
- This reduction is used by both the video coding standards (H.261, MPEG-1, MPEG-2, . . .) and the image coding standards (JPEG).



Spatial frequency

- The human visual system is characterized by a bandpass behavior in the spatial frequency domain.
- In zones where large intensity variations occur, small imperfections are masked (i.e., cannot be seen).





Quality assessment of images

- The techniques for assessing the quality of the images can be classified as subjective or objective.
- A subjective evaluation involves a number of human observers, which can perform absolute or relative assessments.
- In relative assessments, the images are ranked according to the perceived quality.
- In absolute assessments, the observers have to assign a classification, according to some predefined scale, such as:
  - 5. Excellent
  - 4. Good
  - 3. Fair
  - 2. Poor
  - 1. Very poor
- The use of objective criteria (mathematical models) is frequently unavoidable.
- Typically, the objective criteria are based on the mean squared error or on some other similar measures.
- One of the most popular is the peak signal to noise ratio.

Emax: 32; PSNR: 18.5 dB Original Emax: 123; PSNR: 23.9 dB