BASICS ON DIGITAL MULTIMEDIA REPRESENTATION

Fernando Pereira
Instituto Superior Técnico
Multimedia regards content and technologies dealing with a combination of different content forms/modalities, e.g. speech, audio, text, video, images, 3D models, etc.
The Analogue World: Signals
An Analogue World ...

An analog/analogue signal is any variable signal, continuous in both time and amplitude.

* Any information may be conveyed by an analogue signal; often such a signal is a measured response to changes in physical phenomena, such as sound or light, and is obtained using a transducer, e.g. camera or microphone.

* A disadvantage of analogue representation is that any system has noise—that is, random variations—in it; as the signal is transmitted over long distances, these random variations may become dominant.
In modern multimedia, there are many types of relevant signals, also called *media* or *modalities*, used to produce sensory effects, and richer user experiences, notably:

- Text
- Speech
- Audio (includes music)
- Monochromatic and colour imaging
- Monochromatic and colour video
- 3D synthetic models
- Olfactory data
- Haptic data
- …
An audio signal is a representation of sound, typically as an electrical voltage.

Audio signals have frequencies in the audio frequency range of roughly 20 to 20000 Hz (the limits of the human auditory system).

Audio signals may be synthesized directly, or may originate at a transducer such as a microphone, musical instrument pickup, phonograph cartridge, or tape head. Loudspeakers or headphones convert an electrical audio signal into sound.

Audio signals may be characterized by parameters such as their bandwidth, power level in decibels (dB), and voltage level.
The human voice consists of sound made by a human being using the vocal folds for talking, singing, laughing, crying, screaming, etc. Its frequency ranges from about 60 to 7000 Hz.

The human voice is specifically that part of human sound production in which the vocal folds (vocal cords) are the primary sound source.

Generally speaking, the mechanism for generating the human voice can be subdivided into three parts; the lungs, the vocal folds within the larynx, and the articulators, e.g. tongue, palate, cheek, lips.

In telephony, the usable voice frequency band ranges from approximately 300 Hz to 3400 Hz. The bandwidth allocated for a single voice-frequency transmission channel is usually 4 kHz, including guard bands.
Music is an art form whose medium is sound/audio and silence.

The creation, performance, significance, and even the definition of music vary according to culture and social context. Music ranges from strictly organized compositions (and their recreation in performance), through improvisational music to aleatoric forms.

Music can be divided into genres and subgenres, although the dividing lines and relationships between music genres are often subtle, sometimes open to individual interpretation, and occasionally controversial.

The music bandwidth regards the range of audio frequencies which directly influence the fidelity of the music. The higher the audio bandwidth, the better the sound fidelity. The highest practical frequency which the human ear can normally hear is about 20 kHz.

Naturally, music is a very relevant type of audio signal as it is associated to extremely important applications and businesses.
Musical Instruments for all Tastes …
A transducer is a device (commonly implies the use of a sensor/detector) that converts one form of energy to another. Energy types include (but are not limited to) electrical, mechanical, electromagnetic (including light), chemical, acoustic or thermal energy.

A microphone is an acoustic-to-electric transducer that converts sound into an electrical signal.

A loudspeaker is an electroacoustic transducer that produces sound in response to an electrical audio signal input.
An image/video signal is a representation of light, typically as an electrical voltage.

Video corresponds to succession of images at some temporal rate, typically 25 Hz in Europe and 30 Hz in US.

In analogue video, each image/frame is represented as a discrete number of lines, with each line represented by a time-continuous waveform. This means the original 2D continuous signal is converted into a 1D signal using a line by line scanning.

Analogue TV video signals have frequencies in the range of roughly 0 to 5 MHz with this value depending on the image rate and number of lines per image.

Video signals may be synthesized directly, or may originate at a transducer such as a camera. Displays convert an electrical video signal into light.
A transducer is a device (commonly implies the use of a sensor/detector) that converts one form of energy to another. Energy types include (but are not limited to) electrical, mechanical, electromagnetic (including light), chemical, acoustic or thermal energy.

A video camera is an light-to-electric transducer used for image acquisition, initially developed by the television industry but now common in many other applications.

A display is an electric-to-light transducer that produces images in response to an electrical video signal.
Text is the representation of written language which is the representation of a language by means of a writing system.

Text is another form of media corresponding to a sequence of characters that may have to be coded.
Basics on Human Perception

- hearing
- sight
- smell
- touch
- taste
We, the Users …

Audiovisual communication services must, above all, satisfy the final user needs, maximizing the quality of the user experience!
The sensory system for the sense of hearing is the auditory system.

The ability to hear is not found as widely in the animal kingdom as other senses like touch, taste and smell. It is restricted mainly to vertebrates and insects. Within these, mammals and birds have the most highly developed sense of hearing.

- Humans: 20-20000 Hz
- Whales: 20-100000 Hz
- Bats: 1500-100000 Hz
- Fish: 20-3000 Hz
**Physiological Effects: the Thresholds**

- **Threshold of Hearing** — Defines the minimum sound intensity which may be perceived; this threshold varies along the audio band.

- **Threshold of Feeling or Pain** — Defines the sound intensity above which the sounds may cause pain and provoke hearing damages.

Typically, the threshold of pain is about 120 to 140 dB; sound intensity is measured in terms of Sound Pressure Level relatively to a reference intensity with $10^{-16}$ W/cm² at 1 kHz.
Audio Frequency Masking

Auditory masking occurs when the perception of one sound is affected by the presence of another sound.

Auditory masking in the frequency domain is known as simultaneous masking, frequency masking or spectral masking.
The visual system is the part of the central nervous system which enables organisms to process visual detail. It interprets information from visible light to build a representation of the surrounding world.

The visual system accomplishes a number of complex tasks, including the reception of light and the formation of monocular representations, the construction of a binocular perception from a pair of two dimensional projections, the identification and categorization of visual objects, assessing distances to and between objects, and guiding body movements in relation to visual objects.
Binocular vision is vision in which both eyes are used together.

Having two eyes confers at least four advantages over having one:

1. *Gives a creature a spare eye in case one is damaged.*
2. *Gives a wider field of view.* For example, humans have a maximum horizontal field of view of approximately 200 degrees with two eyes, approximately 120 degrees of which makes up the binocular field of view (seen by both eyes) flanked by two uniocular fields (seen by only one eye) of approximately 40 degrees.
3. *Gives binocular summation* in which the ability to detect faint objects is enhanced (the detection threshold for a stimulus is lower with two eyes than with one).
4. *Gives stereopsis* in which parallax provided by the two eyes' different positions on the head give precise depth perception.
While designing a video system, it is essential to account for:

★ The limited human capacity to see spatial detail

★ The conditions under which the human visual system reaches the ‘illusion of motion’

★ The lower sensibility to color in comparison with luminance/brightness
Illusion of Motion: Temporal Resolution

- Visual information corresponds to a time varying 3D signal which has to be transformed into a time varying 1D signal to be transmitted using the available channels.

- At the reception, the information is visualized in a 2D space resulting from the projection (during acquisition) into the camera plane.

- The 2D signal is sampled in time at a rate that guarantees the illusion of motion; this illusion improves with the image rate.

Experience shows that it is possible to get a good illusion of motion up from 16-18 image/s, depending on the image content.

For TV, the frame rate is 25 Hz (Europe) and 30 Hz (US and Japan) due to the electromagnetic interference with the electric network at 50/60 Hz for the old CRT (cathode ray tube) displays.
**Visual Acuity versus Number of Lines**

* Visual acuity regards the eye capability of distinguishing (resolving) spatial detail; it is measured with the help of special test images called *Foucault bars images*.

* The visual acuity determines the minimum number of lines in the image in order the user located at a certain distance does not ‘see’ the lines and gains the sensation of spatial continuity.

* The maximum number of lines that the Human Visual System manages to distinguish in a Foucault bars image is given by

\[ N_{\text{max}} \sim 3400 \frac{h}{d_{\text{obs}}} \]

for \( d_{\text{obs}}/h \sim 8 \), \( N_{\text{max}} \sim 425 \) lines; \( d_{\text{obs}}/h \sim 3 \), \( N_{\text{max}} \sim 1150 \) lines.

Multimedia Communication, Fernando Pereira, 2012/2013
Visual Signal Representation
Black and White versus Colour

- **Black and white (monochrome) imaging** requires the representation of a single signal called *luminance* which indicates *how much luminous power will be detected by an eye looking at the surface from a particular angle of view*. Luminance is thus an indicator of how bright the surface will appear. In this case, the solid angle of interest is the solid angle subtended by the eye's pupil.

- For **colour imaging** visually acceptable results, it is necessary (and almost sufficient) to provide three samples (color channels) for each pixel, which are interpreted as coordinates in some *color space*. The RGB color space is commonly used in displays, but other spaces such as YCbCr, HSV, and are often used in other contexts.
Luminance is a photometric measure of the luminous intensity per unit area of light travelling in a given direction. It describes the amount of light that passes through or is emitted from a particular area, and falls within a given solid angle.

- The **luminous flux** radiated by a luminous source with a power spectrum $G(\lambda)$ is given by:

  $$\Phi = k \int G(\lambda) \, y(\lambda) \, d\lambda \quad [\text{lm or lumen}] \quad \text{with } k=680 \, \text{lm/W}$$

  where $y(\lambda)$ is the average sensibility function of the human eye

- The way the radiated power is distributed by the various directions is given by the **luminous intensity**:

  $$J_L = \frac{d\Phi}{d\Omega} \quad [\text{lm/sr or vela (cd)}]$$

- For video systems, the relevant quantity is the **luminance** of a surface element $dS$ when it is observed with an angle $\theta$ such that the surface orthogonal to the observation direction is $dS_n$

  $$Y = \frac{dJ_L}{dS_n} \quad [\text{lm/sr/m}^2]$$

  which corresponds to the **luminous flux, per solid angle, per unit of area**.
**Rod cells**, or rods (*bastonetes*), are photoreceptor cells in the retina of the eye that can function in less intense light than can the other type of photoreceptor, the **cone cells**. Named for their cylindrical shape, rods are concentrated at the outer edges of the retina and are **used in peripheral vision**. There are about 90 million rod cells in the human retina. More sensitive than cone cells (100 times more), **rod cells are sensitive to luminance** and are almost entirely responsible for **night vision**.

**Cone cells**, or cones, are photoreceptor cells in the retina of the eye that **function best in relatively bright light**. The cone cells gradually become sparser towards the periphery of the retina (there are about 4-6 million in the human eye). Cones are less sensitive to light than the rod cells in the retina (which support vision at low light levels), but **allow the perception of color**. They are also able to perceive finer detail and more rapid changes in images, because their response times to stimuli are faster than those of rods.

Because **humans usually have three kinds of cones** with different response curves and, thus, respond to variation in color in different ways, they have **trichromatic vision**.

As approximately 800000 nerve fibers perform the transport of visual information from the eyes to the brain (about 150 less than the number of receptors), there is some compression mechanism working at the retina: only the differences between adjacent receptors cells are transmitted.
“Colour is a property of the mind and not of the objects in the world; it results from the interaction of a light source, an object, and the visual system.” Newton

Colorimetry studies show that it is possible to reproduce a high number of colours through the addition of only 3 primary colours, carefully chosen.

The primary colours used in most cameras and displays to generate most of the other colours are

- Vermelho (RED)
- Verde (Green)
- Azul (Blue)

Luminance, Y, may be obtained from the primary colours as

\[ Y = 0.3 \, R + 0.59 \, G + 0.11 \, B \]
Chromaticity is an objective specification of a color regardless of its luminance, that is, as determined by its hue and saturation.
Multimedia Communication, Fernando Pereira, 2012/2013
Luminance and 2 Chrominances ...

Y = 0.30R + 0.59G + 0.11B
B - Y = U
R - Y = V

~ 5 MHz
~ 1-2 MHz
~ 1-2 MHz

Y - Luminance  B - Y = U  R - Y = V
Why YUV and not RGB?

YUV is a color space representing a color image or video

1. **Taking human perception into account to allow reduced bandwidth for chrominance components**

2. **Typically enabling transmission errors or compression artifacts to be more efficiently masked by the human perception than using a "direct" RGB-representation.**

While other color spaces have similar properties, an additional reason to adopt YUV would be for better interfacing analog and digital television and also photographic equipment that conform to certain YUV standards.

\[
\begin{bmatrix}
Y' \\
U \\
V
\end{bmatrix} = \begin{bmatrix}
0.299 & 0.587 & 0.114 \\
-0.14713 & -0.28886 & 0.436 \\
0.615 & -0.51499 & -0.10001
\end{bmatrix} \begin{bmatrix}
R \\
G \\
B
\end{bmatrix}
\]
Acquisition, Transmission and Synthesis Signals ...
The Analogue World: Systems
Main Analogue AV Systems

- **Telephone** - The telephone is a telecommunications device that transmits and receives sounds, usually the *human voice*. Telephones are a point-to-point communication system whose most basic function is to allow two people separated by large distances to talk to each other.

- **Radio** - Radio broadcasting is a *one-way* wireless transmission of audio (notably music) signals over radio waves intended to reach a wide audience. Stations can be linked in radio networks to broadcast a common radio format, either in broadcast syndication or simulcast or both.

- **Television** - Television (TV) is a telecommunication medium for transmitting and receiving moving images that can be monochrome (black-and-white) or colored, with accompanying sound. "Television" may also refer specifically to a television set, television programming, or television transmission.
Analogue TV Systems

- **Monochrome** – Only the luminance signal is transmitted; systems with a different number of lines per frame have existed.

- **Colour** – Three signals – *luminance plus two chrominance signals* – are transmitted; systems with a different number of lines per frame exist.

- National Television System Committee (NTSC)
- Phase Alternate Line (PAL)
- Séquentiel couleur à mémoire (SECAM)

![Map of TV systems](image)
The Starting of Analogue TV ...
Portuguese TV Milestones

- 1957 – Start of black and white emission with one RTP channel.
- 1968 – Start of the emissions for the second channel, RTP2.
- 1972 – Start of RTP Madeira.
- 1975 – Start of RTP Açores.
- 1980 – Start of regular colour TV emissions.
- 1992 – Start of SIC emissions, the first private TV channel.
- 1993 – Start of TVI emissions, the second private TV channel.
- 1994 – Start of cable TV.
- 2012 – Switch off of the analogue emissions and start of digital TV emissions with DVB-T.
From Analogue to Digital
Digitization

Process of expressing analogue data in digital form.

Analogue data implies ‘continuity’ while digital data is concerned with discrete states, e.g. symbols, digits.

Vantages of digitization:
★ Easier to process
★ Easier to compress
★ Easier to multiplex
★ Easier to protect
★ Lower powers
★ ...

Multimedia Communication, Fernando Pereira, 2012/2013
Sampling or Time Discretization

Sampling is the process of obtaining a periodic sequence of samples to represent an analogue signal.

Sampling is governed by the Sampling Theorem which states that:

An analog signal may be fully reconstructed from a periodic sequence of samples if the sampling frequency is, at least, twice the maximum frequency present in the signal.
Image Sampling

The number of samples (resolution) of an image is very important to determine the ‘final fidelity/quality’.

The required resolution must take into account at least the human visual system and the final display conditions.
Quantization is the process in which the continuous range of values of a sampled input analogue signal is divided into non-overlapping subranges, and to each subrange a discrete value of the output is uniquely assigned.
2 Levels Quantization

Reconstruction levels

Output values

Input values

Decision thresholds

1 bit/sample image (bilevel)

8 bit/sample image

Multimedia Communication, Fernando Pereira, 2012/2013
4 Levels Quantization

- Input values: 0, 64, 128, 192, 255
- Output values: 0, 32, 96, 160, 224, 255
- Decision thresholds: 64, 128, 192

2 bit/sample image: 8 levels
8 bit/sample image: 256 levels

Reconstruction levels
Uniform Quantization

4 bit/sample
0000, 0001, 0010, 0011, ...

2 bit/sample
00, 01, 10, 11

3 bit/sample
000, 001, 010, 011, 100, 101, 110, 111

1 bit/sample
0, 1
Digitization: the Signal ‘Behind the Bars’ ...

Amplitude

Quantization step

Sampling period

Analogue signal

Sampled and quantized signal

Time
Non-Uniform Quantization

For many signals, e.g., speech, uniform or linear quantization is not a good solution in terms of minimizing the mean square error (and thus the Signal to Quantization noise Ratio, SQR) due to the non-uniform statistics of the signal.

Also to get a certain SQR, lower quantization steps have to be used for lower signal amplitudes and vice-versa.
Pulse Code Modulation (PCM)

PCM is the simplest form of digital source representation/coding where each sample is independently represented with the same number of bits.

- Example 1: Image with 200×100 samples at 8 bit/sample takes $200 \times 100 \times 8 = 160000$ bits with PCM coding

- Example 2: 11 kHz bandwidth audio at 8 bit/sample takes $11000 \times 2 \times 8 = 176 \text{kbit/s}$ with PCM coding

Being the simplest form of coding, as well as the least efficient, PCM is typically taken as the reference/benchmark coding method to evaluate the performance of more powerful (source) coding algorithms.
Image, Samples and Bits ...

Luminance =

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>87</td>
<td>89</td>
<td>101</td>
<td>106</td>
<td>118</td>
<td>130</td>
<td>142</td>
<td>155</td>
</tr>
<tr>
<td>85</td>
<td>91</td>
<td>101</td>
<td>105</td>
<td>116</td>
<td>129</td>
<td>135</td>
<td>149</td>
</tr>
<tr>
<td>86</td>
<td>92</td>
<td>96</td>
<td>105</td>
<td>112</td>
<td>128</td>
<td>131</td>
<td>144</td>
</tr>
<tr>
<td>92</td>
<td>88</td>
<td>102</td>
<td>101</td>
<td>116</td>
<td>129</td>
<td>135</td>
<td>147</td>
</tr>
<tr>
<td>88</td>
<td>94</td>
<td>94</td>
<td>98</td>
<td>113</td>
<td>122</td>
<td>130</td>
<td>139</td>
</tr>
<tr>
<td>88</td>
<td>95</td>
<td>98</td>
<td>97</td>
<td>113</td>
<td>119</td>
<td>133</td>
<td>141</td>
</tr>
<tr>
<td>92</td>
<td>99</td>
<td>98</td>
<td>106</td>
<td>107</td>
<td>118</td>
<td>135</td>
<td>145</td>
</tr>
<tr>
<td>89</td>
<td>95</td>
<td>98</td>
<td>107</td>
<td>104</td>
<td>112</td>
<td>130</td>
<td>144</td>
</tr>
</tbody>
</table>

Binary representation
8 bit/sample -> 256 (2^8) levels

87 = 0101 0111
130 = 1000 0010
Sample - A sample refers to a value at a point in time and/or space. A sampler is a subsystem or operation that extracts samples from a continuous signal. In video, there are luminance and chrominance samples, most of the times not with the same density/size.

Pixel - A pixel is generally thought of as the smallest single element of a digital image (including all components). The more pixels are used to represent an image, the closer the result can resemble the original. The number of pixels in an image is sometimes called the spatial resolution.

- If all the image components have the same resolution, the number of pixels in the image is the number of samples of each component.
- However, if the various components have different resolutions, than the number of pixels corresponds to the number of samples of the component with the highest resolution, typically the luminance.
ITU-R 601 Recommendation: a Typical Spatial Resolution

- Most important standard PCM video/image format
- Basic sampling rate: 13.5 MHz for the luminance and 6.75 MHz for the chrominances
- Considers 625 and 525 lines systems for 25 and 30 Hz systems, respectively (576 and 480 useful lines for 25 and 30 Hz)
- Considers both 4:3 and 16:9 aspect ratios
- Quantization: 8 bit/sample
**Colour Subsampling Formats**

<table>
<thead>
<tr>
<th>Format</th>
<th>Resolution Y</th>
<th>Resolution U/V</th>
<th>Horizontal</th>
<th>Vertical</th>
</tr>
</thead>
<tbody>
<tr>
<td>4:4:4</td>
<td>720 x 576</td>
<td>720 x 576</td>
<td>1:1</td>
<td>1:1</td>
</tr>
<tr>
<td>4:2:2</td>
<td>720 x 576</td>
<td>360 x 576</td>
<td>2:1</td>
<td>1:1</td>
</tr>
<tr>
<td>4:2:0</td>
<td>720 x 576</td>
<td>360 x 288</td>
<td>2:1</td>
<td>2:1</td>
</tr>
<tr>
<td>4:1:1</td>
<td>720 x 576</td>
<td>180 x 576</td>
<td>4:1</td>
<td>1:1</td>
</tr>
<tr>
<td>4:1:0</td>
<td>720 x 576</td>
<td>180 x 144</td>
<td>4:1</td>
<td>4:1</td>
</tr>
</tbody>
</table>
The chroma sub-sampling is generally expressed as a three part ratio J:A:B, describing the number of luminance and chrominance samples in a determined area.

This area has J pixels wide and 2 pixels high, being referred to as *conceptual area*. The value of A defines the number of chrominance samples, CB and CR, in the first row, while B is the number of chrominance samples in the second row of the conceptual area.
Colour Subsampling Targets

- **4:4:4** – Luminance and each chrominance with the same number of samples; targets high quality, professional applications, studios, etc.

- **4:2:2** – Luminance with twice the samples of each chrominance (chrominances with same number of lines but half the sample per line); targets average quality applications such as digital TV and DVD.

- **4:2:0** – Luminance with 4 times the samples of each chrominance (chrominances with half the number of lines and half the samples per line); targets lower quality applications and lower resource systems, notably video in mobile networks and Internet.
**Progressive format** - Progressive scan differs from interlaced scan in that the image is displayed on a screen by scanning each line (or row of pixels) in a sequential order rather than an alternate order, as done with interlaced scanning.

**Interlaced format** - Interlacing divides the lines in a single frame into odd and even lines and then alternately refreshes them at 25/30 frames per second, leading to the so-called odd and even fields.

In other words, in progressive scan, the image lines (or pixel rows) are scanned in ‘regular’ numerical order (1,2,3) down the screen from top to bottom, instead of in an alternate order (lines or rows 1,3,5, etc... followed by lines or rows 2,4,6).
Compression Arrives ...
Why Compressing?

- **Speech** – e.g. $2 \times 4000$ samples/s with 8 bit/sample – $64000$ bit/s = 64 kbit/s

- **Music** – e.g. $2 \times 22000$ samples/s with 16 bit/sample – $704000$ bit/s = 704 kbit/s

- **Standard Video** – e.g. $(576 \times 720 + 2 \times 576 \times 360) \times 25$ (20736000) samples/s with 8 bit/sample – $166000000$ bit/s = 166 Mbit/s

- **Full HD 1080p** - $(1080 \times 1920 + 2 \times 1080 \times 960) \times 25$ (103680000) samples/s with 8 bit/sample – $829440000$ bit/s = 830 Mbit/s
How Much is Enough?

- Recommendation ITU-R 601: 25 images/s with 720×576 luminance samples and 360×576 samples for each chrominance with 8 bit/sample
  \[
  (720 \times 576) + 2 \times (360 \times 576) \times 8 \times 25 = 166 \text{ Mbit/s}
  \]

- Acceptable rate, p.e. using H.264/MPEG-4 AVC: 2 Mbit/s

  => Compression Factor: \(166/2 \approx 80\)

The difference between the resources requested by compressed and non-compressed formats may lead to the emergence or not of new industries, e.g., DVD, digital TV.
Digital Source Coding/Compression

Process through which a source, e.g., images, audio, video, is digitally represented considering relevant requirements such as compression efficiency, error resilience, random access, complexity, etc.

- Example 1: Maximizing the quality for the available rate
- Example 2: Minimizing the rate for a target quality
Source Coding: Original Data, Symbols and Bits

The encoder represents the original digital data (PCM) as a sequence of symbols, and later bits, using in the best way the set of available coding tools, to satisfy the relevant requirements.

The encoder extracts from the original data ‘its best’ ...
Coding … and Decoding …

Encoder

Decoder

PCM Bits

Compressed bits

Much less bits !!!
Digital Image Coding: Main Types

★ LOSSLESS (exact) CODING – The image is coded preserving all the information present in the digital image; this means the original and decoded images are mathematically the same.

★ LOSSY CODING – The image is coded without preserving all the information present in the digital image; this means the original and decoder images are mathematically different although they may still be subjectively the same (transparent coding).

Multimedia Communication, Fernando Pereira, 2012/2013
Where does Compression come from?

★ REDUNDANCY – Regards the similarities, correlation and predictability of samples and symbols corresponding to the image/audio/video data.

-> redundancy reduction does not involve any information loss this means it is a reversible process -> lossless coding

★ IRRELEVANCY – Regards the part of the information which is imperceptible for the visual or auditory human systems.

-> irrelevancy reduction is an irreversible process -> lossy coding

Source coding exploits these two concepts: for that, it is necessary to know the source statistics and the human visual/auditory systems characteristics.
It is essential to keep in mind that visual information is to be consumed by the Human Visual System!

The Human Visual System is the client that must be satisfied in terms of visual quality!
It is essential to keep in mind that audio/speech information is to be consumed by the Human Auditory System!

The Human Auditory System is the client that must be satisfied in terms of audio quality!
Where does Compression come from?

- **REDUNDANCY** – Regards the similarities, correlation and predictability of samples and symbols corresponding to the image/audio/video data.

  -> redundancy reduction does not involve any information loss this means it is a reversible process –> *lossless coding*

- **IRRELEVANCY** – Regards the part of the information which is imperceptible for the visual or auditory human systems.

  -> irrelevancy reduction is an irreversible process –> *lossy coding*

Source coding exploits these two concepts: for that, it is necessary to know the source statistics and the human visual/auditory systems characteristics.
Metadata: Data about the Data
Content, Content, and More Content … How to Get what is Needed?

- Increasing availability of multimedia information
- Difficult to find, select, filter, manage AV content
- Because the value of content depends on how easy it is to find, select, manage and use it!
- More and more situations where it is necessary to have ‘information about the content’
Described content is content structured and accompanied by information called *metadata* allowing its simpler and more adequate search and adaptation to different consumption contexts.
Metadata: Data about the Data

* Content description or metadata regards all types of data features which may be relevant for a more efficient searching, filtering, adaptation, management and, in general, consumption of data.

* Metadata or "data about the data" may:
  - Describe the data/content itself, e.g. genre
  - Describe the data/content coding format, coded quality, etc
  - Describe conditions about the data/content, e.g. licensing
  - ...

The more it is known about the data (metadata), the better the data can be processed, filtered, segmented, coded, adapted, ...
Filtering TV ...
Managing iPods Data ...

- Can't Get It Out of My Head
  Greatest Hits - Electric Light Orchestra

- The Things We Do For Love
  Greatest Hits 1972-76 - 10cc

- I Wouldn't Want To Be Like You
  I Robot - Alan Parsons Project

- Rikki Don't Lose That Number
  A Decade Of Steely Dan - Steely Dan

- Take the Long Way Home
  The Very Best Of Supertramp - Supertramp

- Nights on Broadway
  Their Greatest Hits -The Record (Disk 1) - Bee...

- Your Mama Don't Dance
  The Best Of Friends - Loggins & Messina
YouTube: Metadata, Searching …

YouTube considers metadata fields such as

- Title
- Description
- Category
  - Autos & Vehicles, Comedy, Education, Entertainment, Film & Animation, Gaming, Howto & Style, Music, News & Politics, People & Blogs, Pets & Animals, Science & Technology, Sports, Travel & Events, …
- Date of upload
- Number of views
- Scores
- …
Content Description Features

- **Features conveying information that cannot be deduced from the content:**
  - e.g. recording date & conditions, author, copyright data, conditions for access, context of recording, coding format, links to other relevant material

- **Features conveying information that is present in the content (explicit content):**
  - **LOW LEVEL** – metadata allowing automatic extraction but without (direct) semantic value
  - **HIGH LEVEL** – metadata related to the human interpretation of the content (semantics), typically with manual or at most semi-automatic extraction
Looking for Napoleon or looking for ‘Napoleon’?
Speaking a More Natural Language?

Columbia University, NY.
Speaking a new ‘Audio Language’ …

The Police: Roxanne

DB
Search by Example: Simple but Powerful ...
“I want to see the goals of a Benfica-Sporting ...”

“Football goal” low-level definition =

- Fast horizontal movement +
- Cheering from the crowd +
- Slow motion repetition +
- Fixed shot of the centre of the field
Defining Quality ...
Quality: a Simple yet Difficult Concept

Like many human sensations, quality is easy to understand but difficult to define.

According to the Wikipedia:

- A quality (from Latin - qualitas) is an attribute or a property.
- Some philosophers assert that a quality cannot be defined.
- In contemporary philosophy, the idea of qualities and especially how to distinguish certain kinds of qualities from one another remains controversial.
Quality is like an Elephant …

*The blind men and the elephant: Poem by John Godfrey Saxe*
Quality of Service (QoS) refers to a collection of networking technologies and measurement tools that allow for the network to guarantee delivering predictable results.

- **Quality of Service (QoS)**
  - Resource reservation control mechanisms
  - Ability to provide different priority to different applications, users, or data flows
  - Guarantee a certain level of performance (quality) to a data flow

- **(Service) Provider-centric concept**
Changing Landscape …

- YouTube
- Facebook
- Twitter
- Blu-ray Disc
- Flickr
- HDTV
- Multimedia Communication, Fernando Pereira, 2012/2013
Multimedia is about sharing experiences (real or imaginary) with others …

In a way, it all started with story telling and wall drawing around the fire in the caves of early men …

Modern multimedia systems are evolved versions of the good old story telling and wall drawing, which hopefully offer increasingly richer experience.

Recognizing the central role of the user, the degree of richness of the (individual) experience may be measured by the so-called Quality of Experience (QoE) …
Quality of Service versus Quality of Experience

- **Quality of Service** - Value of the average user’s service richness estimated by a service/product/content provider

- **Quality of Experience** - Value (estimated or actually measured) of a specific user’s experience richness

Quality of Experience is the dual (and extended) view of Quality of Service

\[
\text{QoS} = \text{provider-centric} \\
\text{QoE} = \text{user-centric}
\]
Factors Impacting Quality of Experience

- Technical factors
  - Task Application
  - Social and psychological factors
- Device
- Network
- Content Format
- User
- Environment
- Content
- User Expectations
- Context
How Shall a Multimedia Experience Be?

Depending on the specific application, it may have to be:

- **Faithful** - accuracy
- **Truthful** – realistic if relevant, synchronization
- **Individual** – emotional
- **Immersive** – natural, multimodal consistency
- **Contextual** - adaptive
- **Engaging** – fun, intense
- **Effective** – fast, recognition
- **Useful** – task performing
- **Interactive** – natural, short delay
- **Intuitive, Easy** – interfaces
- **...**
Usability is QoE…
Usability is QoE …
QoE modeling may consider more or less quality dimensions depending on the service/product, each with a different weight on the overall assessment.

User centered quality dimensions are expected to be dominating.
A QoE Model Example: IPTV

- Video quality
- Audio quality
- Audiovisual synchronisation
- Error resilience
- Random access
- Channel hopping delay
- Interface usability
- Navigation capabilities
- Personalization capabilities
- Metadata quality
- Immersion
- …
Compression Performance Assessment ...
Compression Metrics

Compression Factor = \frac{\text{Number of bits for the original PCM image}}{\text{Number of bits for the coded image}}

Bit/pixel = \frac{\text{Number of bits for the coded image}}{\text{Number of pixels in the image (typically Y samples)}}

The number of pixels of an image corresponds to the number of samples of its component with the highest resolution, typically the luminance.
Quality Metrics

Subjective evaluation

Objective evaluation

\[ \text{MSE} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (y_{ij} - x_{ij})^2 \]

\[ \text{PSNR(dB)} = 10 \log_{10} \frac{255^2}{\text{MSE}} \]

\( X(m,n) \rightarrow \text{Compression} \rightarrow \ Y(m,n) \)

e.g., scores in a 5 levels scale

\( x \) and \( y \) are the original and decoded data

Multimedia Communication, Fernando Pereira, 2012/2013
And, finally, Transmission ...
**Channel Types**

- Data transmission, digital transmission, or digital communications is the physical transfer of data (a digital bit stream) over a point-to-point or point-to-multipoint communication channel.

- There are so-called ‘**guided**’ channels and ‘**atmospheric**’ channels depending if some form of cable or the atmosphere are used for the transmission. Examples of such channels are copper wires, optical fibres, wireless communication channels, and storage media.

- The data are represented as an electromagnetic signal, such as an electrical voltage, radiowave, microwave, or infrared signal.

- While analog transmission is the transfer of a continuously varying analog signal, digital communications is the transfer of discrete messages.
Typical Digital Transmission Chain ...

Source

Digitalization
(sampling + quantization + PCM)

Source Coding

Channel Coding

Modulation

Channel

Analog signal

PCM bits

Compressed bits

‘Channel Protected’ bits

Modulated symbols

Multimedia Communication, Fernando Pereira, 2012/2013
Channel coding is the process applied to the bits produced by the source encoder to increase its robustness against channel or storage errors.

- At the sender, redundancy is added to the source compressed signal in order to allow the channel decoder to detect and correct channel errors.

- The introduction of redundancy results in an increase of the amount of data to transmit. The selection of the channel coding solution must consider the type of channel, and thus the error characteristics, and the modulation.

Block Codes

Symbols with useful information | Correcting symbols
---|---
m | k
n
\[ R = \frac{m}{n} = 1 - \frac{k}{n} \]
Baseband versus Modulated Transmission

Baseband Transmission

- In telecommunications, baseband refers to signals and systems whose range of frequencies is measured from close to 0 hertz to a cut-off frequency, a maximum bandwidth or highest signal frequency.
- Baseband can often be considered a synonym to lowpass or non-modulated, and antonym to passband, bandpass, carrier-modulated or radio frequency (RF).

Modulated Transmission

- In telecommunications, modulation is the process of conveying a message signal, for example a digital bit stream or an analog audio signal, inside another signal that can be physically transmitted.
- Modulation varies one or more properties of a high-frequency periodic waveform, called the carrier signal, with a modulating signal which typically contains information to be transmitted.
- Modulation of a sine waveform is used to transform a baseband message signal into a passband signal.
Line Coding

* In telecommunication, a line code (also called digital baseband transmission method) is a code chosen for use within a communications system for baseband transmission purposes.

* Line coding consists in representing the digital signal to be transported by an amplitude- and time-discrete signal that is optimally tuned for the specific properties of the physical channel (and of the receiving equipment).

* The waveform pattern of voltage or current used to represent the 1s and 0s of a digital data on a transmission link is called *line encoding*. The common types of line encoding are unipolar, polar, bipolar, and Manchester encoding.
Modulation is the process through which one or more properties of a carrier (amplitude, frequency or phase) vary as a function of the modulating signal (the signal to be transmitted).

Any of these properties can be modified in accordance with a low frequency signal to obtain the modulated signal.

The selection of an adequate modulation is essential for the efficient usage of the bandwidth of any channel.

Together, (source and channel) coding and modulation determine the bandwidth necessary for the transmission of a certain signal.
Factors to consider in selecting a modulation:

- Channel characteristics
- Spectrum efficiency
- Resilience to channel distortions
- Resilience to transmitter and receiver imperfections
- Minimization of protection requirements against interferences

Basic digital modulation techniques:

- Amplitude modulation (ASK)
- Frequency modulation (FSK)
- Phase modulation (PSK)
- Mix of phase and amplitude modulation (QAM)
For 64-QAM, only 64 modulated symbols are possible!
Digital TV: a Full Example

- ITU-R 601 Recommendation: 25 images/s with 720×576 luminance samples and 360×576 samples for each chrominance with 8 bit/sample
  \[ (720 \times 576) + 2 \times (360 \times 576) \times 8 \times 25 = 166 \text{ Mbit/s} \]

- Acceptable rate after source coding/compression, p.e. using H.264/AVC: 2 Mbit/s

- Rate after 10% of channel coding 2 Mbit/s + 200 kbit/s = 2.2 Mbit/s

- Bandwidth for video information in a digital TV channel, e.g. with 64-PSK or 64-QAM: 2.2 Mbit/s / \( \log_2 64 \) \( \approx 370 \text{ kHz} \)

- Number of digital TV channels / analogue TV RF slot: 8 MHz / 400 kHz \( \approx 20 \text{ channels} \)
Typical Digital Transmission Chain...

Source

- Digitalization (sampling + quantization + PCM)
- Analog signal
- PCM bits
- Source Coding
- Compressed bits

Channel

- Channel Coding
- ‘Channel Protected’ bits
- Modulated symbols
- Modulation

Multimedia Communication, Fernando Pereira, 2012/2013
Bibliography

