Audio & Video Compression and its Application in Consumer Products
Introduction - The evolution of Audio/Video consumer products and the role of compression techniques.

Audio & Video compression principles

Audio compression

Video compression

Audio/Video synchronisation

The MPEG model and its situation in a communication context

Application to DVD (Digital Versatile Disc)

Application to DVB (Digital Video Broadcasting)

Conclusion
Agenda

操纵

Introduction - The evolution of Audio/Video consumer products and the role of compression techniques.

- Audio & Video compression principles
- Audio compression
- Video compression
- Audio/Video synchronisation
- The MPEG model and its situation in a communication context
- Application to DVD (Digital Versatile Disc)
- Application to DVB (Digital Video Broadcasting)
- Conclusion
The evolution of our products (1/5)
The evolution of our products (2/5)

- The STB (Set-Top-Box) as the link between the home and the world-wide information infrastructure.
The evolution of our products (3/5)

- The STB (in home) as the gateway to various services.
  Local Server provides 2 kind of services:
  
  * Broadcast
    Analogue & digital TV, NVOD, PPV
  * Point-to-point (Home to local server)
    Home shopping, VOD, e-mail, Web browsing, PC connection...

![Diagram of network setup with local server and internet connection to up to 800 homes.](image-url)
The evolution of our products (4/5)

- The STB as a key element of the home network

![Home Network Diagram]

- Residential Gateway
- Computer
- Television
- DVD Jukebox
- Disk Recorder
- To telephone Network
- To satellite Network
- To cable Network
The evolution of our products (5/5)

- 3C Convergence - Progressive
- New products combine all 3 functions
- Products always more and more complex
- Working in group becomes more and more important
- Philips international co-operation
  (avoiding to reinvent the wheel, concentration on core competencies)
- Products have always new features
- Lifetime of products is always shorter

- The DVD will also be provided with the communication feature (IEEE1394 bus) for home bus integration.
Factors enabling such evolution

Compression is one among the various factors that enable multimedia:

- Audio/Video Compression (e.g. CCIR601 vs MPEG)
- Disc capacity (DVD), communication going digital (Modems, xDSL, ATM ...)
- Electronics (Memory capacity, clock frequency, µP architecture, ... --> decoding at low cost)
- Software (methodology, user interface ...)
- International cooperation (interoperability & economy of scale)
Agenda

- Introduction - The evolution of Audio/Video consumer products and the role of compression techniques.
- Audio & Video compression principles
  - Audio compression
  - Video compression
  - Audio/Video synchronisation
  - The MPEG model and its situation in a communication context
  - Application to DVD (Digital Versatile Disc)
  - Application to DVB (Digital Video Broadcasting)
- Conclusion
Compression in first A/V Products (1)

- First Audio/Video products make compression without knowing it was compression.
- How?
  - By removal of irrelevancies
- Audio and Video characteristics

<table>
<thead>
<tr>
<th></th>
<th>Audio</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral Sensitivity</td>
<td>Good</td>
<td>Bad</td>
</tr>
<tr>
<td>Spatial Sensitivity</td>
<td>Bad</td>
<td>Good</td>
</tr>
</tbody>
</table>
Compression in first A/V Products (2)

- Audio products
  From 2 to 7.1 channels are enough to provide the spatial resolution.

- Video products
  Three colours (RGB) are enough to provide the spectral resolution.
The need for more compression (1/5)

- Audio: Compression needed in spectral domain
- Bitrate of a stereo audio source (CD-DA encoding)

Sampling frequency: 44.1 kHz
Stereo
16-bit per sample
Bitrate = 44100 * 2 * 16 = 1.41 Mbit/sec
The need for more compression (2/5)

- Video: Compression needed in spatial domain
- Bitrate of a video source (CCIR 601 - 50 Hz countries)

![Video image diagram]

25 images per second
YUV coding (Y: luminance - U,V : Chrominance)
Y: 8 bit per pixel - U,V: 1 pixel on 2 coded, 8 bit per pixel
Bitrate = \((576 \times 720) \times 25 \times 16 = 166 \text{ Mbit/sec}\)
The need for more compression (3/5)

- Channels availables for AV transmission
  - Analog television channel (compatibility)
    Cable (bandwidth = 8 MHz)
    Satellite (Bandwidth = 30-40 MHz)
  ⇒ Capacity around 40 Mbit/sec

- Compact disc (CD)
  For 74 min. play time : 1.41 Mbit/sec
The need for more compression (4/5)

- MPEG-1 target
  (Video-CD : 74 min. constraints)

[Diagram showing compression rates: Video: 166 Mbit/sec, Audio: 1.4 Mbit/sec]

But quality was judged too poor (about VHS quality)
The need for more compression (5/5)

- MPEG-2 target
  - Program stream (DVD)
    - 1 program (video, multichannel audio, ....)
      - Compression
        - 3-9 Mbit/sec (variable bitrate)
          - but higher quality than MPEG-1
        - = motivation for the capacity increase of the CD (--> DVD)
  - Transport stream (DVB)
    - n programs (video, multichannel audio, ....)
      - Compression
        - about 40 Mbit/sec (constant bitrate)
          - (DVB-Satellite & DVB-Cable)
Principles of compression (1/2)

Compression (or source coding) is achieved by suppressing information:

- redundant information
- irrelevant information

Suppression of redundant information

⇒ lossless compression

example: PCM to DPCM, DCT

The original signal and the one obtained after encoding and decoding are identical
Principles of compression (2/2)

- Suppression of irrelevant information
  ⇒ lossy compression
  Example: bandwidth limitation, masking in audio

The original signal and the one obtained after encoding and decoding are different but are perceived as identical
Audio Demonstration

From “Borderline” Madonna - Stereo - 16 bit/channel
Compression used AAC

- Compression
  - 705 kbps
- Decompression
  - 128 kbps
  - 64 kbps
  - 32 kbps
  - 16 kbps
MOS scale (1/2)

- Signal distortion is not a good measure of the performance of a lossy compression method
  ⇒ another method is necessary: MOS scale (Mean Opinion Score)

- The five-grade CCIR impairment scale (Rec.562)
  1 (Very annoying), 2 (Annoying), 3 (Slightly annoying), 4 (Perceptible but not annoying), 5 (Imperceptible)

- Example: Double blind test
MOS scale (2/2)

Listener answers to:
1. Which signal is the original? 2 or 3?
2. Grade the other one?

Original signal

Compressed signal

Exchange box
(Random for listener)

Selector
(controlled by listener)

Listener answers to: 1. Which signal is the original? 2 or 3? 2. Grade the other one?

Impairment scale

Compressed

Original

Max value

Mean value

Min value
Compression to VBR or CBR

- **CBR (Constant Bit Rate) vs VBR (Variable Bit Rate)**
- **Scene more complex ⇒ Higher bit rate for same quality**
- **CBR ⇒ variable quality (example: Video CD artefact)**
- **Constant quality ⇒ VBR necessary (e.g.: DVD-Video)**
Video demonstration
The compression trade-off

- Compression techniques are still making progress
- Trade-off Complexity/Quality/Bit Rate
- New technique may result in new trade-off
Agenda

- Introduction - The evolution of Audio/Video consumer products and the role of compression techniques.
- Audio & Video compression principles
  - Audio compression
- Video compression
- Audio/Video synchronisation
- The MPEG model and its situation in a communication context
- Application to DVD (Digital Versatile Disc)
- Application to DVB (Digital Video Broadcasting)
- Conclusion
Audio compression in MPEG (1/5)

- Based on psycho-acoustics
- Compress the bit rate without affecting the quality perceived by the human ears (based on the imperfection of human ears)
- Removal of irrelevancies
- 4 main principles:
  - Threshold of audibility
  - Frequency masking
  - Critical bands
  - Temporal masking
Audio compression in MPEG (2/5)

- Principle 1: Threshold of audibility
  - Not all frequency components need to be encoded with the same resolution. \( Nr_{\text{bit}}(f) = \frac{(\text{signal/threshold})_{\text{db}}}{6} \)
Audio compression in MPEG (3/5)

- Principle 2: Frequency masking
  ⇒ Analysis of the incoming signal

![Graph showing frequency masking effect]

- Sound level (db) vs. Frequency (kHz)
- New threshold of audibility due to masker signal
- Masker signal
- Masked signal
Audio compression in MPEG (4/5)

Principle 3: Critical bands

* Human ear may be modelled as a collection of narrow band filters
* Bandwidth of these filters = critical band
* Critical band
  (<100 Hz) for lowest audible frequencies
  (∼4 kHz) for highest audible frequencies
* The human ear cannot distinguish between two sounds having two different frequencies in a critical band.
  Example: when we hear 50 & 100 Hz at the same time we cannot distinguish them.
* Consequence:
  Noise masking threshold depends solely of the signal energy within a limited bandwidth domain.
  The largest sound is taken as the representative of the critical band.
  Necessity to analyse the signal at 100Hz resolution at low-frequency
Audio compression in MPEG (5/5)

- Principle 4: temporal masking
  ⇒ selection of the frame duration for frequency analysis and encoding.

![Diagram showing temporal masking principles](chart)

- Pre masking
- Envelope of masker
- Post masking
- Simultaneous masking

Time ~100msec

Level (db)

~(-10msec) 0 0 ~100msec
An enabling tool: the filter bank (1/2)

Digital audio signal, Fs

Analysis

Filter bank
n subband signals, Fs
DECIMATION
n subband signals, Fs/n
ENCODING / DECODING
n subband signals, Fs/n
INTERPOLATION
n subband signals, Fs
Synthesis

Filter bank
Reconstructed signal, Fs
An enabling tool: the filter bank (2/2)

- After decimation, same bit rate as original signal, but signal decomposed in various frequency ranges ⇒ possibility of frequency based compression
- Filter-bank:
  Aliasing occurs due to decimation
- It exists a class of filter-bank such that aliasing is compensated in synthesis filter: QMF (Quadrature Mirror Filter) but high complexity
- Pseudo-QMF (Polyphase filter bank) is used. Has good compromise between computation cost and performances
- Remark: Aliasing may occur if signal in a adjacent band is not reconstructed with an adequate resolution.
The MPEG encoder

Digital audio input

Filter Bank

Subband samples

Bit allocation

Quantised samples

Bitstream formatting

Encoded bitstream

Psychoacoustic model

Signal to mask level for each subband (quantisation information)

Ancillary data
The MPEG filter bank

- In MPEG, 32 equal-width subbands are used
- For each subband, necessity to define the maximum signal level and the minimum mask level.
- **BUT, at low frequencies:**
  bandwidth of subbands > critical bands
- ⇒ Necessity to rely on an FFT in order to compensate the lack of frequency selectivity of filterbank at low frequencies
Psychoacoustic model & Bit allocation(1/2)

- An FFT compensates the lack of frequency selectivity of filterbank at low frequencies
- FFT: 512 samples (layer 1) & 1024 samples (layer 2)
  resolution for layer 1: $\frac{F_s}{512} < 100$ Hz
- A psychoacoustic model based on the FFT computes the signal to mask ratio for each subband (1 bit = 6db)
- Ideally, after allocation, quantisation noise < masking level
- The scale factors are computed for each subband from the filterbank output (floating point representation of samples)
- The bit allocator adjust the bit allocation in order to meet the bitrate requirement.
- The bitstream syntax is dependent of the MPEG layer (See later)
Psychoacoustic model & Bit allocation (2/2)

Frequency

Signal/Mask
Level (db)

Level (db)

Mask level

Signal level

Bit allocation

= 1 bit

Frequency
The MPEG decoder

- Decoder is simple (Complexity is at encoder side)
- Remark 1: DCC is MPEG-1 but DCC encoder has no FFT, relies only on power in the 32 subbands
  ⇒ Higher bit rate (320 kbps) to reach transparent quality
- Remark 2: MPEG specifies bitstream syntax only. Encoder are given for information. Possibility of improvement.
Audio features in MPEG

❖ **MPEG1**:
  * Mono/stereo/dual/joint stereo ( Possibility Dolby surround )
  * Sampling frequencies : 32, 44.1 & 48 kHz
  * 3 layers : trade-off complexity/delay versus coding efficiency of compression
  * Various bit rate : trade-off quality versus bitrate

❖ **MPEG2**:
  * 5.1 channels
  * Sampling frequencies extended to 16, 22.05 & 24 kHz
Dolby surround principles (1/5)

- 4 channels carried by stereo pair \(\Rightarrow\) same tools as for stereo
- Compatible with stereo installation

**Configuration**

**From 4 channels to 2 channels**

- Left (L)
- Center (C)
- Surround (S)
- Right (R)

```
+  - 3 dB  +
  |       |
  |       |
Center (C)  BP  Surround (S)
  |       |
  |       |
-90°      100 Hz - 7000 Hz +90°
  |       |
  |       |
Right (R)  Phase shifter  Right (R)
```
## Dolby surround principles (2/5)

### Encoding equations

- \( \text{Lt} := L + \frac{1}{\sqrt{2}} \cdot C + \frac{1}{\sqrt{2}} \cdot S \)
- \( \text{Rt} := R + \frac{1}{\sqrt{2}} \cdot C - \frac{1}{\sqrt{2}} \cdot S \)
- \( \text{C} := \frac{\text{Rt} + \text{Lt}}{\sqrt{2}} \)
- \( \text{S} := \frac{\text{Lt} - \text{Rt}}{\sqrt{2} \cdot j} \)

### Simple decoder

- \( \text{L} := \text{Lt} \quad \Rightarrow \quad \text{Ld} := L + \frac{1}{\sqrt{2}} \cdot C + \frac{1}{\sqrt{2}} \cdot S \)
- \( \text{R} := \text{Rt} \quad \Rightarrow \quad \text{Rd} := R + \frac{1}{\sqrt{2}} \cdot C - \frac{1}{\sqrt{2}} \cdot S \)
- \( \text{Cd} := \frac{1}{\sqrt{2}} \cdot \text{L} + \frac{1}{\sqrt{2}} \cdot \text{R} + \text{C} \)
- \( \text{Sd} := -j \cdot \frac{1}{\sqrt{2}} \cdot \text{L} - j \cdot \frac{1}{\sqrt{2}} \cdot \text{R} + \text{S} \)
Dolby surround principles (3/5)

- Simple decoder provides only 3 dB channel separation
  (See previous equations)
  ⇒ Need for improvement
  ⇒ Dolby Surround pro-logic decoder (next slide)
Dolby surround principles (4/5)

Dolby surround pro-logic decoder

![Diagram of Dolby surround pro-logic decoder]

- Lt
- Rt
- Ld
- Cd
- Sd
- Rd
- L'
- C'
- S'
- R'

Direction Compensation

VCA

-3 dB

BP

-90°
Dolby surround principles (5/5)

Performance of Dolby pro-logic decoder
Channel separation larger than 35 dB
5.1 surround sound

MPEG-2 surround configurations (front/back)

- 3/2
- 3/0 + 2/0
- 3/1
- 2/2
- 2/0 + 2/0
- 3/0
- 2/1
- 2/0
- 1/0

+ LFE (opt.)

(Fs/96)

15-120Hz
Virtualisation

- Virtualisation has no direct relation with the MPEG standard. It is considered here only because it may be implemented in some of the future audio products (DVD, STB ...)
- Virtualisation is a product feature.
- It allows reproduction of surround information (5.1, 3/1) on a stereo installation.
Virtualisation principle

- Virtualisation = processing of the signal in such a way the source of the signal is perceived at a selected position outside the loudspeaker axis (virtual loudspeaker).

- Drawback : very sensitive to listener position (stability)

- Remark : a mono signal coded in normal stereo is perceived between the two loudspeakers
Stereo widening

- Also called Q-sound™, incredible sound, azimuth positioning...
- The stereo sources are positioned at virtual locations for improving the stereo effect (cheap analog solution exists)
- Real sound comes from real loudspeakers. Perceived sound is as if stereo signals were coming from virtual loudspeakers
Virtual surround

- Virtual surround gives on a stereo installation the subjective effect of a multichannel configuration.

- Each channel is virtually positioned at a location around the listener.

The stereo installation performs the addition of the processed signals for each audio channel.

- Real sound comes from a stereo installation. Perceived sound is as if the various surround signals were coming from some virtually located loudspeakers.
# Summary of surround aspects

<table>
<thead>
<tr>
<th>Stereo source</th>
<th>Stereo receiver</th>
<th>Multichannel receiver</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereo source</td>
<td>Stereo widening (incredible sound, azimuth positioning)</td>
<td>3-D special effects (hall, theatre, stadium)</td>
</tr>
<tr>
<td>Multichannel source</td>
<td>Virtual surround</td>
<td>Multichannel output</td>
</tr>
</tbody>
</table>

**Remarks about Dolby surround pro-logic:**
- Only carrier is stereo, source & presentation are multichannel
- Compatible with stereo installation (no surround effect except in the case of surround virtualisation)
Agenda

✦ Introduction - The evolution of Audio/Video consumer products and the role of compression techniques.
✦ Audio & Video compression principles
✦ Audio compression
✦ Video compression
✦ Audio/Video synchronisation
✦ The MPEG model and its situation in a communication context
✦ Application to DVD (Digital Versatile Disc)
✦ Application to DVB (Digital Video Broadcasting)
✦ Conclusion
Video compression in MPEG (1/6)

- **Principles**
  
  - **removal of intrapicture redundancy:**
    Image is decomposed in 8*8 pixels subimages. Each subimage contains redundant information.
    DCT transformation (in frequency domain) decorrelates the input signal. (most energy in low spatial frequencies)
  
  - **removal of interpicture redundancy:**
    Coding of difference with an interpolated picture (moving vectors)
  
  - **high frequent spatial frequencies quantized with lower resolution than low ones** (remove irrelevancy)
  
  - **zig-zag scan and VLC** (remove redundancy)
Video compression in MPEG (2/6)

- Result
  - 4:2:2 CCIR 601 resolution: 166 Mbps
    \(=25\text{images/sec} \times 576\text{lines} \times 720\text{pixels} \times 2(\text{lum & chrom}) \times 8\text{bits}\)
    \(\Rightarrow \pm 3-4\text{ Mbps (mean) in MPEG2}\)
  - 4:2:0 SIF resolution: 30 Mbps
    \(=25\text{images/sec} \times 288\text{lines} \times 352\text{pixels} \times 1.5(\text{lum & chrom}) \times 8\text{bits}\)
    \(\Rightarrow \pm 1.2\text{ Mbps (CBR) in video CD (MPEG1)}\)
Video compression in MPEG (3/6)

- Spatial redundancy reduction (DCT example)

```
139 144 149 153 155 155 155 155
144 151 153 156 159 156 156 156
150 155 160 163 158 156 156 156
159 161 162 160 159 159 159 159
159 160 161 162 162 155 155 155
161 161 161 161 160 157 157 157
162 162 161 163 162 157 157 157
162 162 161 161 163 158 158 158
```

```
1260  -1 -12 -5  2  -2  -3   1
-23  -17  -6  -3  -3  0  0  -1
-11  -9  -2  2  0  -1 -1  0
-7  -2  0  1  1  0  0  0
-1  -1  1  2  0 -1  1  1
 2  0  2  0  -1  1  1 -1
-1  0  0  -1  0  2  1  -1
-3  2  -4  -2  2  1  -1  0
```

DCT

```
158  0  -1  -1  0  0  0  0
-1  0  0  0  0  0  0  0
1  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0
0  0  0  0  0  0  0  0
```

Quantisation

```
158  0  -1  -1  -1  -1  EOB
```

zig-zag scan
Video compression in MPEG (4/6)

- Temporal redundancy reduction

<table>
<thead>
<tr>
<th>Order of presentation</th>
<th>Increase of compression rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 I B B P B B P B I B</td>
<td>I: Intra-coded picture</td>
</tr>
<tr>
<td>1 B B P B B P B B P I</td>
<td>P: Predicted picture</td>
</tr>
<tr>
<td>2 B P B B P B B P I B</td>
<td>B: Bi-directionally interpolated picture</td>
</tr>
</tbody>
</table>

Bi-directional prediction

<table>
<thead>
<tr>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 B B B P B I B</td>
</tr>
<tr>
<td>1 B B B P B I B</td>
</tr>
<tr>
<td>2 B B B P B I B</td>
</tr>
<tr>
<td>3 B B B P B I B</td>
</tr>
</tbody>
</table>

Order of transmission:

| 0 I P B B P B B B |
| 1 B B P B B B B P |
| 2 B B P B B B B P |
| 3 B B P B B B B P |

Increase of compression rate:

| 0 I B B P B B P B I B |
| 1 B B P B B P B B P I B |
| 2 B P B B P B B P I B |
| 3 B B B P B I B |
| 4 B B B B B B B B B |
| 5 B B B B B B B B B |
| 6 B B B B B B B B B |
| 7 B B B B B B B B B |
| 8 B B B B B B B B B |
| 9 B B B B B B B B B |
Video compression in MPEG (5/6)

- Model of a possible encoder
Video compression in MPEG (6/6)

✦ MPEG1 en MPEG2 video features

* MPEG1
  ✦ sequential picture
  ✦ resolution: SIF format 288(240)*356*24,25 or 30 Hz

* MPEG2
  ✦ sequential or interlaced
  ✦ various levels: low level (SIF: 288*356), main level (CCIR601: 576 * 720), high 1440 level (HDTV: 1152*1440), high level (EQT: 1152*1920)
  ✦ various profiles (toolboxes): simple profile (No B picture), main profile (=MPEG1+interlaced), SNR scalable profile (allows graceful degradation (noise improvement at same resolution), spatial scalable profile (hierarchical coding: improvement at higher resolution), high profile.
Agenda

- Introduction - The evolution of Audio/Video consumer products and the role of compression techniques.
- Audio & Video compression principles
- Audio compression
- Video compression
- Audio/Video synchronisation
- The MPEG model and its situation in a communication context
- Application to DVD (Digital Versatile Disc)
- Application to DVB (Digital Video Broadcasting)
- Conclusion
Synchronisation

Synchronisation in the multimedia context refers to the mechanism that ensures a temporal consistent presentation of the audio-visual information to the user.
Intramedia synchronisation

\[ \Delta T = \text{Constant} \]

\[ \Delta T_1 \]

\[ \Delta T_2 = \Delta T_1 \]

\( \Delta T \) between capture & presentation = Constant

≡ Same clock frequency & Data on time

⇒ Need for corresponding tools
Intermedia synchronisation

\[ \Delta T_{\text{Audio}} = \Delta T_{\text{Video}} \]
\[ \equiv (\text{Sampled at the same time} \Rightarrow \text{Presented at the same time}) \]
\[ \Rightarrow \text{Possible tools: common time base and presentation control (media synchronisation with the common time base)} \]

Ex.: Lip_sync (requirement: \(|\text{delay}\_\text{difference}| < 80\text{msec}\)
Recovery of clock in CBR

- CBR = Constant Bit Rate
- if the clock to recover is synchronous with transport clock ⇒ Recovery of clock but not of common time base
- Remark: possibility to slave DSM to local clock

Diagram:
- CBR stream
- Time Information carried by each sample
- Filling level 50%
- Phase error
- Filter
- VCO
- Recovered clock
- Processing

Equation:
\[ \text{CBR} = \text{Constant Bit Rate} \]
Recovery of clock and time base in VBR

- **VBR = Variable Bit Rate**
- **Need for insertion of time stamps (OUTPUT TIME)**
  Output time stamp says for example: “It is now 16h25”
  Receiver adjusts its own horloge to the received time stamp
- **Recovery of clock & of common time base**

![Time Stamp Diagram]

1. **Clock**
2. **Data stream**
3. **Time stamps**
4. **Time stamp extraction**
5. **Channel**
6. **Recovered clock**
7. **Error**
8. **Filter**
9. **VCO**

Time information carried only by time stamps
Synchronisation with common time base

- Insertion of time stamp (=INPUT TIME)
  Input time stamp says: “Sample has been sampled at 16h29”.
  Receiver presents the sample at (its input time stamp + maximum encoding and decoding delay).
  Alternative: transmission of presentation time stamp (input time+delay)
Getting data on time

- "On time" ⇒ Not too late, not too early
  No buffer over- or underflow
- Flow control: not applicable in broadcasting
- Common time base and
  Definition of a standard target decoder that describes the data consumption pattern of the receiver.
- Remark: Direct MPEG (microsoft) does not use time information for clock recovery but relies on flow control
Streams

- Idea of continuity (pipelining)
- Carry time information for clock recovery
- No flow control (allows broadcasting)
  The emitter must have a precise knowledge of the receiver data consumption pattern (explicit in MPEG STD)
- Just-in-time
  Shorter delay and smaller buffer size than with flow control
- Two aspects in synchronisation:
  Clock recovery & timing control (model & buffering)
Requirement on the channel for stream transport

Data information ⇒ BER (Bit Error Rate) requirement
No repetition of frame possible ⇒ FEC (Forward Error Correction)

Time information ⇒ No jitter
Agenda

- Introduction - The evolution of Audio/Video consumer products and the role of compression techniques.
- Audio & Video compression principles
- Audio compression
- Video compression
- Audio/Video synchronisation
- The MPEG model and its situation in a communication context
- Application to DVD (Digital Versatile Disc)
- Application to DVB (Digital Video Broadcasting)
- Conclusion
What is MPEG? (1/2)

- Moving Picture Expert Group
- International standard (ISO/IEC)
  ⇒ Interoperability & economy of scale
- Compression of audio and video and multiplexing in a single stream
- Definition of the interface not of the codecs
  ⇒ room for improvement
- MPEG-1: until 1.5 Mbps, for DSM
  Progressive picture, stereo (Dolby surround)
What is MPEG ? (2/2)

- **MPEG-2**: Various bit rates (CBR & VBR)
  Program stream for DSM, transport stream for network
  Interlaced picture, 5.1 audio channels
  Definition of various video levels (e.g. CCIR601 resolution: 4-9 Mbps, HDTV: 15-25 Mbps) and profiles
- **MPEG-3**: Cancelled, integrated in MPEG-2
  (Initially: for HDTV)
- **MPEG-4**: Standard for audio, video and graphics in interactive
  2D and 3D multimedia communication.
  (Initially: low bit rate for real-time personal communication)
- **MPEG-7**: Multimedia contents description interface
- **MPEG-21**: Focus on multimedia distribution and on DRM aspects.
The MPEG model (1/2)

[Diagram showing the MPEG model with components labeled as follows:
- Audio signal
- Video signal
- Captured signals
- Presented signals
- Audio encoder
- Video encoder
- Multiplexer
- Transmission channel
- Demultiplexer
- Audio decoder
- Video decoder
- Digital storage medium or Network]
The MPEG model (2/2)

- Compression of audio & video and multiplexing in a single stream
- Guarantees intramedia and intermedia synchronisation.
- MPEG defines an interface
  - bitstream syntax
  - timing of the bitstream $\Rightarrow$ STD specifying timing requirement (ideal model)

- Consequences:
  - Decoder should compensate deviations from STD
  - Network should correct jitter introduced by the channel (RTD-LJ)

- MPEG stream must be adapted to transmission channel formatting, error correction, channel coding (b.v.video-CD)
Components of the MPEG standard

- The MPEG standard is composed of 3 main parts:
  - Audio: Specifies the compression of audio signals
  - Video: Specifies the compression of video signals
  - System: Specifies how the compressed audio and video signals are combined in the multiplexed stream (program stream or transport stream).

- Each part specifies:
  - The bitstream syntax
  - The timing requirement and the related information (bit rate, buffer needs)
Synchronisation Mechanism (1/2)

Audio input
- Assemble audio frames, Sample STC for PTS
- Audio encoder
- Sample STC for PCR (SCR)
- Multiplexer and system encoder
- Transmission channel

Video input
- Assemble pictures, Sample STC for PTS
- Video encoder
- STC

Audio output
- Comparison PTS and STC and presentation
- Audio decoder
- System decoder and demultiplexer

Video output
- Comparison PTS and STC and presentation
- Video decoder
- Extraction of PCR (SCR)
- STC
Synchronisation Mechanism (2/2)

- PCR for TS & SCR for PS (but same concept)
- Clock & time base recovery: Time-stamping at OUTPUT (PCR included in TS multiplex, SCR in pack header)
- Audio & video clock locked to STC ⇒ easy recovery (see next slide)
- Synchronisation of audio & video to common time base (Time stamping at Input)
- STD is defined (because of the absence of flow control) streams are such that STD buffers never over- or underflow
- In TS, many program in a single stream but unique clock per program.
- Time information ⇒ No Jitter for transport
Clock recovery in receiver

- PCR
- Error
- Low Pass Filter (Integrator)
- VCO
- Video clock divider
- Audio clock divider
- Comparison PTS and STC and Presentation
- Audio output
- Video clock
- Audio clock
- Decoded audio
- PTS
- STC
- Load first PCR
- STC (Counter)
- STC (Counter)
MPEG program & transport streams

- **Program streams:**
  - Relatively error free environment
  - program stream packet may have variable and great length
  - Single time base

- **Transport streams:**
  - environment where errors are likely
  - many programs (independent time base)
  - Transport stream packet : fixed, 188 bytes
  - Contains tables
MPEG in a communication context (1)

“Typical” communication system
MPEG in a communication context (2)

- MPEG: Source coding only (bit rate reduction) + multiplexing
- The MPEG stream must be adapted to the channel in what concern its physical characteristics and in order to get the required QoS (Quality of Service) & Security
  - Encryption
  - Channel coding (forward error correction, interleaving, modulation codes)
  - multiplexing & formatting
  - modulation (frequency allocation)
  - multiple access method
- Some channels: CD/DVD - satellite - cable - ATM - 1394
MPEG in a communication context (3)

* A simple view of MPEG in the communication context

![Diagram showing MPEG compression layers and multiplexing for different communication channels like Cable, Satellite, Disc, DVB, DVD, etc.](image-url)
Agenda

- Introduction - The evolution of Audio/Video consumer products and the role of compression techniques.
- Audio & Video compression principles
- Audio compression
- Video compression
- Audio/Video synchronisation
- The MPEG model and its situation in a communication context
- Application to DVD (Digital Versatile Disc)
- Application to DVB (Digital Video Broadcasting)
- Conclusion
CD : Some concepts

- Hard disk vs compact disc : more differences than just storage technique.
  HD developed for data storage and recording, CD developed for stream storage (CD-DA) ⇒ their basic differences

- Questions
  * track form?
  * read direction? Why?
  * CAV or CLV? Why?
  * Access time : CD-ROM vs HD?
  * Data storage: on which face?
  * Production method?
  * Capacity?
  * Sensitivity to error? Diameter of a possible hole?
  * Standard = Interface definition : CD vs HD ?
CD-DA: Encoder model (1/3)

A/D conversion

Left

Right

6 samples
= 24 bytes
= 1 frame

Error correction
encoding

PCM 44.1 kHz
16 bit/sample/channel

32 bytes
/frame

Modulation

EFM + 3 merging bits

Subcode
(1 byte / frame)

Synchronisation pattern
27 bit/frame

561 bits / frame

588 channel bit /frame

Physical layer

CD-ROM
1 sector = 98 frames
75 sectors/sec.
The CD-DA physical layer adapts the input stream (audio) to the requirements of the channel:

- **Modulation**: EFM (Eight to fourteen modulation + 3 merging bits)
  - Pit & land length (number of successive 0 or 1 as written to disc): between 3 and 11 channel bits
  - DC free code for adaptation to the channel bandwidth & for clock recovery considerations.

- **Error correction**: (Cross-interleaved Reed-Solomon code)
  - Interleave placed between C1 & C2 ECC.
  - Next slide presents only principles and not real CD implementation.
CD-DA: Encoder model (3/3)

* Error correction: addition of redundancy in order to be able to correct errors (e.g. RS(28,24,5) * RS(32,28,5))

Principle:

<table>
<thead>
<tr>
<th>24 bytes</th>
<th>4 bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>C1 codewords</td>
</tr>
<tr>
<td>28 bytes</td>
<td></td>
</tr>
<tr>
<td>C2 codewords</td>
<td></td>
</tr>
<tr>
<td>4 bytes</td>
<td></td>
</tr>
</tbody>
</table>

* Interleaving: time diversity in order to deal with error burst.
Successive errored channel bits (burst error) do not damage the same Reed-Solomon table.
CD-ROM encoder model

Sync Pattern 12 bytes

Header 4 bytes
User data 2048 bytes
Zeroes 8 bytes

Header 4 bytes
Subheader 8 bytes
User data 2048 bytes
Subheader 8 bytes
User data 2324 bytes
Optional EDC 4 bytes

Additional error detection & error correction encoding

2340 bytes

Mode 1
OR
Scrambling
2340 bytes

Mode 2
OR

Video-CD uses CD-ROM mode 2 sectors

1 CD-ROM sector = 2352 bytes

For EDC only

Optional EDC 4 bytes

User data 2336 bytes

Additional error detection & error correction encoding

2340 bytes

Sync Pattern 12 bytes

Header 4 bytes
User data 2048 bytes
Zeroes 8 bytes

Header 4 bytes
Subheader 8 bytes
User data 2048 bytes
Subheader 8 bytes
User data 2324 bytes
Optional EDC 4 bytes

Additional error detection & error correction encoding

2340 bytes

Mode 1
OR
Scrambling
2340 bytes

Mode 2
OR

Video-CD uses CD-ROM mode 2 sectors

1 CD-ROM sector = 2352 bytes
The CD standards

- CD-DA 1982
- CD-ROM 1984
- CD-ROM-XA
- CD-i 1987
- CD-bridge
- Photo-CD
- Video-CD 93
- CD-BGM
- CD-MIDI
- CD-G
- CD-i Ready
- CD-WO 90
- CD-RW 96
From CD to DVD: the motivation

- Motivation = increase the capacity

Why? - Requirement of the motion picture industry
- Playback time: more than 135 min. (duration of 90% of films)
- Picture quality: superior to laser disc
- Audio quality: 5.1 channels surround
- Language/subtitles: 3 languages minimum.

⇒ capacity needs: more than 4.7 Gbytes

Where? - In physical layer

DVD: developed specifically for audio/video (≠ video CD).
The DVD physical layer (1/2)

- Recorded sectors: 2366 bytes (13 * 182 bytes)
- Synchronisation: (2*13)*32 channel bits (equiv. to 2418 bytes)
- 8/16 modulation: 37856 channel bits
- Data sector: 2064 bytes (12 * 172 bytes)
- ECC (per group of 16 sectors): 38688 channel bits

- Scrambling:
  - ID (incl. sector#): 4 bytes
  - EDC: 6 bytes
  - CPR-MAI: 6 bytes

- Data: 2048 bytes

- EDC: Error Detection Code
- ECC: Error Correction Code
- CPR-MAI: CoPyRight MAagination Information

- Row Interleaving:
  - Recorded sectors: 2366 bytes (13 * 182 bytes)
The DVD physical layer (2/2)

- Objective was the storage of 2K sectors
- Error Correction Code (Reed-Solomon) - add redundancy

![Diagram showing the DVD physical layer structure]

- Modulation - time diversity
  (Number of consecutive 0 : between 2 and 10)
  Pit and land length : between 3 and 11 (Idem CD)
- Synchronisation : for sector reconstruction.
DVD: the capacity improvement (1/4)

- Increase of channel bit density (gain = 4.50)
  - Min pit length: \((0.83\mu \Rightarrow 0.4\mu)\)
  - Track pitch: \((1.6\mu \Rightarrow 0.74\mu)\)
  - Diameter of laser spot (\(\div\) wavelength/NA)
    - Wavelength (780nm \(\Rightarrow\) 640 nm) \(\Rightarrow\) gain = 1.5
    - NA (0.45 \(\Rightarrow\) 0.60) \(\Rightarrow\) gain = 1.78
    - reduced margin \(\Rightarrow\) gain = 1.68

- Modulation:
  - EFM (8 to 17 bit) \(\Rightarrow\) 8 to 16 \(\Rightarrow\) gain = 1.06

- Error correction
  - RS(32,28,5)*RS(28,24,5) \(\Rightarrow\) RS(182,172,11)*RS(208,192,17) \(\Rightarrow\) gain = 1.16
DVD: the capacity improvement (2/4)

- No subcode ⇒ gain = 1.03
- Sync pattern ⇒ gain = 1.03
- Better sector formatting
  sector length (2352 bytes ⇒ 2064) ⇒ gain = 1.14
- Other (e.g. recorded area) ⇒ gain = 1.07

Total gain : 7.2
Capacity per side : 650 MBytes (mode 1) ⇒ 4.7 Gbytes
DVD: the capacity improvement (3/4)

Single-layer single-sided disc

0.6 mm

Single-layer double-sided disc

B side

0.6 mm

A side

Dual-layer single-sided disc

0.6 mm

For layers 0 and 1

Dual-layer double-sided disc

0.6 mm
DVD: the capacity improvement (4/4)

- Capacity of the various types

  Single-layer single-side        4.7 Gbytes
  Dual-layer single-side         8.5 Gbytes
  Single-layer double-side       9.4 Gbytes
  Dual-layer double-side         17 Gbytes
The 3 components of the DVD-V standard

DVD = DVD (Not “Digital Video Disc”, Not “Digital Versatile Disc”)
DVD-V = DVD - Video

Contents of the data block.
How audio and video are mapped to the block, file and volume structure

How blocks may be retrieved.
Definition of the file and volume structure.

How blocks of 2048 bytes are stored on the disc

Part 3 : Video specification
Part 2 : File system specification
Part 1 : Physical specification
Some DVD-V features (1/2)

Presentation data = MPEG program stream, VBR, max peak bit rate = 10.08 Mbps

<table>
<thead>
<tr>
<th>Video data</th>
<th>1 stream</th>
<th>Mpeg1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mpeg2 (ML@MP)</td>
</tr>
<tr>
<td>Audio data</td>
<td>max 8 streams</td>
<td>Mpeg2 + 7.1 ext.Dolby AC-3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Linear PCM (incl. 96 kHz - 24 bits)</td>
</tr>
<tr>
<td>Sub picture data</td>
<td>max 32 streams</td>
<td>Run length encoded Bit map</td>
</tr>
<tr>
<td>(subtitles)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Some DVD-V features (2/2)

- Seamless playback
  Language
  parental lock
  Multi-angle camera

- System menu
  Audio stream selection
  Subtitle selection
  Angle

- Encryption
  Decryption key hidden on the disc.
The DVD family of products

<table>
<thead>
<tr>
<th></th>
<th>DVD-ROM</th>
<th>DVD-Video</th>
<th>DVD-Audio</th>
<th>DVD-R</th>
<th>DVD-RAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part 3: Application</td>
<td>Video specification</td>
<td>Audio specification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part 2: File system</td>
<td>UDF</td>
<td>UDF &amp; ISO9660</td>
<td>UDF</td>
<td>(UDF &amp; ISO9660)</td>
<td></td>
</tr>
<tr>
<td>Part 1: Physical</td>
<td>Read only</td>
<td>Write-once</td>
<td>Rewritable</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Recording on disk - principle

- CD principle: reflectivity of pits & lands are different. Pits and lands are used to store 0 and 1.
- CD-RW principle: reflectivity of the two phases of the recording material (amorphous, crystalline) are different. Controlling the phase allows storage of 0 or 1.
- To Amorphous state (low reflectivity): T above melting point (600°C) & fast cooling
- To Crystalline state (high reflectivity): T above 200°C for a sufficient time
- Recording: by the laser heating the recording layer
- Reading: by laser as for CD (-> compatibility)
Introduction - The evolution of Audio/Video consumer products and the role of compression techniques.

Audio & Video compression principles

Audio compression

Video compression

Audio/Video synchronisation

The MPEG model and its situation in a communication context

Application to DVD (Digital Versatile Disc)

Application to DVB (Digital Video Broadcasting)

Conclusion
Adaptation to the DVB channel

- **Channel coding**: transforms the TS in an other sequence of bits containing the same information than the input stream but more robust against the imperfections of the transmission on the physical channel.
  - Cost: a higher bit rate

- **Modulation**: transforms an input sequence to an analog waveform for transmission over the physical channel.
Channel coding (1/3)

- Unlike source coding that removes redundancy, channel coding adds redundancy in a structured way so that the decoder be able to detect and/or correct the errors introduced by the physical channel.
Channel coding may include:

- Spectral modification of the signal for adaptation to the channel (e.g. remove DC, spectrum shaping like uniform distribution in the frequency space ...)

- FEC: Forward Error Correction
  Addition of redundancy in order to allow error detection and/or correction (example: The total of bought articles is similar to a parity byte)

![Diagram showing channel coding](image)
Channel coding (3/3)

* Interleaving
Time diversity in order to deal with error bursts. The successive bytes of information are dispersed in time on the transmission channel in such a way that an error burst does not affect neighbouring bytes. Interleaving is often combined with FEC so that error bursts could be corrected by the FEC.

Example:

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ... AEIMBFJNCGKODHLQP....
```

```
AB&DEF&HIJKLM&OPQ...... AEIMBFJ&KODHLQP....
```

```
|---|---|---|---|
```

Interleaving

Deinterleaving

### Channel

```
& : Erroneous byte
| : Beginning of an error correcting block
- : Element of an error correcting block
```

----> A burst of errors affects bytes belonging to different error correction blocks
Modulation in DVB (1/3)

◆ Different modulation techniques:
  * Cable: QAM
  * Satellite: QPSK
  * Terrestrial: OFDM

◆ Why?
Modulation technique depends on:
  * Physical characteristics of the channel
  * Compatibility constraints with actual analog transmission
Modulation in DVB (2/3)

- Example: influence of SNR on modulation technique selected
  ⇒ QPSK for satellite and QAM for cable

![Diagram showing BER vs SNR for different modulation techniques (QPSK, 16QAM, 32-QAM, 64-QAM)]
Modulation in DVB (3/3)

- **Satellite**
  - Bandwidth: generally 27-36 MHz
  - SNR low: about 10 dB (power transmitted by satellite)
  - Direct path

- **Cable**
  - Bandwidth: 8 MHz (50Hz countries) - 6 MHz (60Hz countries)
  - SNR strong: about 25 dB
  - Echoes from impedance mismatch in the network

- **Terrestrial**
  - Bandwidth: idem as cable
  - Multipath interference, signal level variation, ...
From TS to the DVB channel

- Some blocks are identical for all standards (Cable, Satellite & Terrestrial)
- Inner & outer: terminology is derived from the view of the quasi-error-free channel composed of a transmitter and a receiver.
- Satellite & Terrestrial: More sensitive to error ⇒ inner coder is added
Agenda

- Introduction - The evolution of Audio/Video consumer products and the role of compression techniques.
- Audio & Video compression principles
- Audio compression
- Video compression
- Audio/Video synchronisation
- The MPEG model and its situation in a communication context
- Application to DVD (Digital Versatile Disc)
- Application to DVB (Digital Video Broadcasting)
- Conclusion
Questions

?