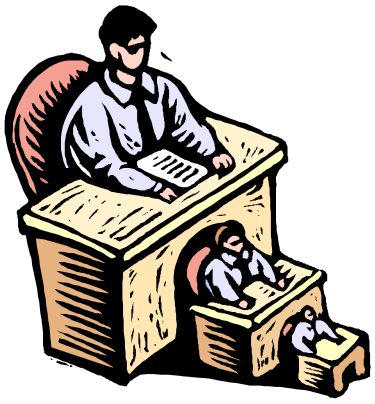
AUDIOVISUAL COMMUNICATIONS

MEEC and MERC



EXERCISES

(with abbreviated solutions)

Fernando Pereira

INSTITUTO SUPERIOR TÉCNICO

Departamento de Engenharia Electrotécnica e de Computadores

February 2010

Exercises for the Audiovisual Communications Course

1. Analogue TV

1.1) 1st Exam 1992/93, 7 June 1993

In March 1980, just a few hours after color TV started in Portugal, two friends (A and B) are discussing at the bar …

A – "Due to backward compatibility, the introduction of color TV does not imply any change in the video signal that monochrome receivers use for the scanning and thus there is no change at all in the video seen by the users in comparison with previous monochrome transmission."

B – "That's not correct! What is true is that the color composite video signal now used only guarantees a reasonable backward compatibility because luminance and chrominances stay in the same bandwidth."

Comment these opinions, notably stating who is right or not and why.

1.2) 2nd Exam 1992/93, 28 June 1993

Consider a high definition analogue TV system with the following characteristics:

- *1250 lines per image, including the vertical retrace*
- *25 images per second*
- *form factor: 16/9*
- *2:1 interlacing*
- *useful duration of a line scanning: 0.99*
- *fraction of lines with video content: 0.98*

a) What is the bandwidth needed for the luminance signal assuming that the visual vertical and horizontal resolutions (in the sense of detail capability) are similar (explain the origin of the equation to be used) ?

b) What is the bandwidth needed for each of the chrominance signals, assuming that the chrominance human visual acuity is half the luminance human visual acuity ?

1.3) 2nd Exam 1993/94, 8 July 1994

As you know, the introduction of color TV (NTSC, PAL and SECAM) was made with the condition of not changing the spectrum allocation which means the bandwidth slot after modulation for each channel had to stay the same. In this context, indicate:

a) Which signals have been selected for color TV

Note: Take as reference for questions a, b and c, the PAL system.

b) What reasons motivated the selection of the signals indicated in a).

c) How was possible to introduce the signals transmitted for color TV in the same total bandwidth before used only by a single signal.

d) Which strategy does the SECAM system use to eliminate the problems regarding color mixtures?

1.4) 1st Exam 1992/93, 7 June 1993

Consider a TV system with the following characteristics:

- *number of 'image elements' per line: 1250*
- *duration of a line: 64 μs*
- *useful duration of a line scanning: 0.92*
- *Kell factor: 0.7*
- *form factor: 16/9*

a) Using only the values above, and without making any specific further assumption, determine the minimum bandwidth (base band) needed for the luminance signal.

b) Assuming that the maximum number of bars in a Foucault image visible in any circumstance in this TV system is 650, state if there is higher visual resolution in the vertical or horizontal direction.

c) Explain in detail which phenomenon justifies the need to introduce the Kell factor in the study of TV systems.

1.5) 1st Exam 1993/94, 22 June 1994

As you may know, the Football World Cup is held every four years. Together with the Olympic Games, this cup is the event most watched in our planet.

a) Discuss which are the possible artifacts, effects, and limitations that a TV system like PAL may have in the faithful reproduction of the T-shirts of the players (assume that all the system behaves in an ideal way, e.g. no channel distortions).

Note: For those less prepared, the T-shirts are typically very colorful and often with vertical or horizontal stripes.

b) Which could be the advantages and disadvantages of seeing a match at your neighbor's TV set who has an NTSC set and receives the matches via satellite?

1.6) 2nd Exam 1998/99, 28 July 1999

Consider a TV system.

a) State which are the characteristics (3) of the human visual system with higher impact in the selection of the spatial and temporal video resolutions to be used in a digital TV system. (R: Limited capability to see spatial detail, limited capability to see temporal detail and thus ability to create the illusion of motion, and lower sensibility to color than to luminance)

b) Explain what motivated the adoption of interlacing in analogue TV systems.

c) Which is the main difference between the PAL and SECAM TV systems on the way color information is transmitted? Which is the main advantage of the SECAM solution?

d) Why do the I and Q signals in the NTSC TV system have different bandwidth?

1.7) 1st Exam 2003/2004, 28 June 2004

Consider an analogue or digital TV system as required.

a) Explain what was the impact on the selection of the signals for colour TV resulting from the need to provide forward compatibility ?

b) The aspect ratio of the TV displays has been changing ... State the aspect ratio values associated to this change – before and after – and explain the motivation for this evolution.

c) Assuming you have at your disposal the typical bandwidth of analogue TV channel, e.g. 8 MHz, determine how many MPEG-2 Video coded digital channels with ITU-R 601 resolution can fit at most in that bandwidth if it is necessary to achieve the quality objectives defined for MPEG-2 Video for secondary distribution and a 64-QAM modulation is used; consider that the audio data uses a rate equal to 1/20 of the video rate

2. Facsimile

2.1) 1st Exam 1993/1994, 22 June 1994

Consider the following extracts of the coding tables for the Modified Huffman Method (MHM) included in the ITU-T T.4 standard.

a) Compute the compression factor for the following sequences of white (W) and black (B) run:

   1) 2W 324B 2W 321B 4W 3B 320W 3B

   2) 2W 3B 2W 4B 4W 2B 1W 3B

b) How do you explain the figures obtained considering the compression factors that are usually expected in a source coding context?

c) Is it possible to include a bit error in the coded bitstream that a group 3 facsimile cannot detect? In which conditions?

White Runs	VLC	Black Runs	VLC
0	00110101	0	0000110111
1	000111	1	010
2	0111	2	11
3	1000	3	10
4	1011	4	011
320	00110110	320	000000110011
384	00110111	384	000000110100
448	01100100	448	000000110101

2.2) 2nd Exam 1992/1993, 7 July 1993

Indicate the MAIN difference between:

a) Group 1 and group 2 facsimiles

b) Group 2 and group 3 facsimiles

c) Group 1 and group 4 facsimiles

d) Group 3 and group 4 facsimiles

2.3) 1st Exam 1992/1993, 7 June 1993

Consider the Modified READ Method (MRM) for the coding of facsimile images.

a) Indicate which coding modes are available and explain in detail the motivation for the definition of each of them.

b) Explain for which reason the group 4 facsimiles can, and even should, use the Modified Modified READ Method (MRMM), this means MRM coding with k= ∞, and may also not used EOF (End of Line) codewords, what may only happen for group 3 facsimiles in very especial conditions ?

2.4) 2nd Exam 1993/1994, 20 July 1994

Consider a facsimile transmission system.

a) Indicate which characteristic would you analyze if you were asked to verify if the codewords in a VLC table were generated by the Huffman method as entropy encoder ?

b) Indicate which coding modes are available in MRM coding and explain in detail the motivation for the definition of each of them.

c) Indicate two techniques used in MRM coding to limit the propagation of errors in a facsimile decoded image.

2.5) 2nd Exam 1993/1994, 8 July 1994

Consider a group 3 facsimile system.

a) Knowing that the used bitrate is 14400 bit/s and that the facsimile modem uses a QAM modulation carrying 6 bit/symbol, compute which is the minimum necessary bandwidth for the signal.

b) Explain for which reason all the phases of the communication protocol, with the exception of the message transmission and call setup phases, use HDLC framing.

c) Indicate for which reason the MHM make-up codes must always be followed by a terminating code even in the case that a run multiple of 64 is to be transmitted.

2.6) 2nd Exam 1998/1999, 28 July 1999

Consider a group 3 facsimile transmission, at 3200 bit/s, for pages with 1000 lines, each line with 1728 samples; assume also that, on average, each line has 80% of white samples.

a) Indicate the constraint for k in order that the global compression factor is higher than 30. Assume that the unidimensionally coded lines have a compression factor of 10 and 20 for the black and white runs, respectively, and the bidimensionally coded lines have, on average,

compression factors which are twice the compression factors for the unidimensionally coded lines.

b) Indicate how much would be the increase in transmission time regarding the previous situation (using the minimum value for k), if a k value which is at least half the previous value is now used to increase error resilience.

c) For the k value in b), how much is the periodicity of unidimensional line coding refreshment in bits ? Which are the advantages and disadvantages (2) of increasing the k value ?

2.7) 1st Exam 2005/2006, 20 June 2006

Consider a facsimile transmission using the Modified READ method for pages with 2500 lines each with 1728 pixels. Assume that, on average, each line has 80% of white pixels and the transmission is 4800 bit/s.

a) Assuming that, for MLST = 0 ms, the unidimensionally coded lines have a compression factor of 10 for the black lengths, that this compression is half the compression achieved for the white lengths, and bidimensional encoding reaches compression factors which are 70% higher regarding unidimensional compression, determine the global compression factor for MLST = 20 ms and k=4 if it is known that 10% of the lines in the image (whatever the type of coding) need a number of bits which is lower than the minimum required, more precisely 20% lower than that minimum.

b) In the conditions above and for MLST = 0 ms, determine which would be the global compression factors for an image if only unidimensionally or bidimensionally (with the exception of the first) coded lines are used.

c) Determine which are the mathematical expressions relating the transmission time with the transmission rate for a set of 25 facsimile pages, if only unidimensionally or bidimensionally (with the exception of the first) coded lines are used and MLST = 0 ms.

3. Digital Imaging

3.1) 2nd Exam 1993/1994, 20 July 1994

Consider the transmission of digital images with a resolution of 720×576 luminance samples and half this resolution, in each direction, for the chrominances (when used), using a channel at 2 Mbit/s.

a) Considering the transmission channel is available during 10 s, how many complete bi-level images are you able to transmit without any compression ?

b) And how many complete grey images can you transmit, in the same 10 s, if 128 grey levels are available (again without compression) ?

c) Considering now that a compression algorithm with compression factors of 20 and 15 for the luminance and chrominances, respectively, is used at 7 bit/sample, how many complete images can you transmit still in the same 10 s ?

3.2) 2nd Exam 1992/1993, 7 July 1993

a) As you know, some of the compression algorithms adopted by the JPEG standard are based on transform coding. Which are the main requirements for a transform to be used in the context of an image compression system ? Why ?

b) Considering the DCT is a linear transform, explain for which reason does Recommendation H.261 computes the transform of the residuals instead of the difference of the transforms if they are mathematically equivalent.

c) Which is the main reason justifying the Lohscheller matrices to be different for the luminance and chrominances ? How does that fact impact the values in the matrices ?

3.3) 1st Exam 1992/1993, 7 June 1993

Consider the JPEG compression standard for multilevel digital images.

a) Identify and describe the various operation modes specified in this standard.

b) For which reasons does the standard specify all these operation modes ?

3.4) 2nd Exam 1993/1994, 8 July 1994

Consider the JPEG compression standard for digital images.

a) Compute the total transmission time, using a 64 kbit/s channel, in the sequential mode, for an image in ITU-R 601 format (720×576 luminance samples and 360×576 samples for each chrominance with 8 bit/sample), considering that the compression factors are 15 and 20 for the luminance and chrominances, respectively.

b) Considering now the transmission to be made using the hierarchical mode, compute the transmission time for the 3 layers used assuming that:

- *The transmission channel is the same.*
- *The spatial resolution for the base layer is 360×288 luminance samples and 180×288 samples for each chrominance.*
- *The spatial resolution doubles, in both directions, for each new layer.*
- *Akways 8 bit/samples.*
- *The compression factors for each layer (and thus for its spatial resolution) increase 25 % for each new layer regarding the previous layer.*
- *The compression factors for the base layer are those indicated above.*

c) State the relative benefits and drawbacks of using the two modes mentioned above, notably considering the transmission times computed.

3.5) 2nd Exam 2005/2006, 8 July 2006

a) Determine the average number of bits per pixel (considering the luminance and chrominances) used when a 4:2:0 image with 8 bit/sample is coded with a global (luminance and chrominances) compression factor of 16. Determine the same metric if a compression factor of 20 is used for the luminance and a compression factor of 12 for the chrominances.

b) What is the main difference between a lossless and a lossy image coding system ? Which of the two types of systems is typically more important ? Why ?

c) State a normative and a non-normative impact in terms of JPEG image compression from the fact that the human visual system is less sensitive to the higher frequencies than to the lower frequencies.

d) Why is entropy coding used in most source coding systems, including JPEG codecs ? What is the greatest disadvantage of entropy coding for transmissions in mobile environments ?

3.6) 1st Exam 2005/2006, 20 June 2006

Assume that a user wants to access a database with JPEG coded images to find some images. The maximum spatial resolution is 720×576 for the luminance and 360×576 for the chrominances, both with 8 bit/sample.

a) Determine which JPEG coding modes have been used to code the images in the database if it is known that the users may access, in an efficient way, versions of the same image in several qualities and spatial resolutions.

b) Assume that the average compression factor for the luminance and chrominances is 10 and 15, respectively, for sequential coding. Assume also that, for the base layer of the progressive coding mode, the compression factors are twice as high and that, for the next layer, the compression factors are 3 times higher than for the sequential coding mode. Knowing that, on average, each user quickly browses 4 images in the base layer before finding the target image, that the user takes 2 s to decide if an image is the target image or not, and that the transmission is made at 64 kbit/s, determine the total time required, on average, to get an image in the maximum quality if images with only images coded with 2 layers are used and that all image layers are sequentially transmited unless the user stops this by selecting the next image. Finally, assume that the decoding times are negligible.

c) Identify which would be the main consequences (at least 2) if the JPEG standard would have used a spatial transform with base functions not independent from the image to code.

3.7) 1st Exam 2008/2009, 24 June 2009

Consider the JPEG and JPEG 2000 image coding standards.

a) State 3 advantages of the type of transform used in the JPEG 2000 standard regarding the transform used in the JPEG standard.

b) For what reason does the JPEG 2000 standard define two RGB to YCrCb transformations ?

c) What type of information is successively sent in the JPEG 2000 standard to obtain spatial resolution OU quality scalability for a certain target resolution ?

3.8) 2nd Exam 2008/2009, 20 July 2009

Consider the JPEG and JPEG 2000 image coding standards.

a) Explain (and justify) how many (2D) DWT decompositions must be applied to obtain a JPEG 2000 bitstream with 4 spatial resolutions.

b) Explain (and justify) how many (2D) DWT decompositions must be applied to an image with spatial resolution 1200×1200 to obtain a JPEG 2000 bitstream making available a decoded image with a spatial resolution of 75×75.

c) Explain on what process mainly depends the maximum quality obtained for each of the 4 resolutions above (in c).

d) Explain what is the objective of coding each bitplane of each band using three coding passages.

4. Videotelephony, Videoconference and H.261

4.1) 1st Exam 1992/1993, 7 June 1993

Consider the specification of a video compression algorithm for digital videotelephony based on the 4 main tools used in ITU-T Recommendation H.261. Assume the video data has a 10 Hz frame rate and a 360×288 luminance spatial resolution; chrominance spatial resolution is half the luminance spatial resolution in both directions.

a) Considering that each luminance and chrominance sample is represented with 8 bits, compute the bitrate without compression needed for the transmission of this video data.

b) To limit error propagation along the various frames, 2 coding modes are created, depending on the coding tools used. From the 4 main available tools mentioned above, state and justify which tools can be used for each coding mode while achieving the identified target with the best coding performance.

c) Compute the global compression factor assuming that for each image all the coding tools mentioned above are used with the following compression factors. Assume that the compression factors are independent in the sense that their effects may be accumulated.

Coding Tool	Luminance Compression Factor	Chrominance Compression Factor
1	5	5
2	3	4
3	2	1
4	1	1

4.2) 1st Exam 1993/1994, 22 June 1994

As you know, ITU-T Recommendation H.261 is an international standard for the compression of video data in videotelephony and videoconference applications.

a) State for which reason the estimation and compensation of motion is made at macroblock level (16×16 pixels).

b) State the reason why the motion vectors are coded in a differential way.

c) Explain the motivation for the fact that no DCT coefficients selection thresholds are specified for the quantization process. Which is the main advantage of this option ?

d) Explain the motivation for the fact that DC DCT coefficients are quantized differently from AC DCT coefficients. What is the difference ?

4.3) 2nd Exam 1993/1994, 8 July 1994

Consider the video compression algorithm for videotelephony and videoconference specified in the ITU-T Recommendation H.261. For a certain video sequences, the probabilities for the various macroblock (MB) coding classes were measured and the results in the table below were obtained.

a) Indicate the set of codewords to code each MB coding class if Huffman entropy coding is used.

b) Indicate 3 reasons that may justify the high percentage of macroblocks coded with the Intra mode.

c) Which factors determine the choice of the maximum amplitude for the motion vectors components ?

Coding Type	**Coding Mode**	**Probability**
A	Intra	0,25
B	Inter	0,6
C	Inter + Motion Compensation	0,05
D	Inter + Motion Compensation + Filter	0.05
E	Inter + Filter	0,05

4.4) 2nd Exam 1993/1994, 20 July 1994

Consider a videotelephony communication using ITU-T Recommendation H.261 with a bitrate of 64 kbit/s. The video data is coded using a CIF spatial resolution at 10 Hz.

a) Knowing that 12800, 3200 and 32000 bits were spent on the coding of the first, second and third images, respectively, compute (justifying) which are the minimum acceptable size for the encoder output buffer and the initial visualization delay at the receiver, considering that the sequences of images mentioned above includes the worst case in terms of bit production.

b) Indicate, justifying, which is the maximum size of the encoder output buffer if a maximum initial visualization delay of 200 ms is required.

4.5) 1st Exam 1994/1995, 30 June 1995

Consider the ITU-T Recommendation H.261 for the coding of video data in videotelephony and videoconference applications.

a) Indicate the 3 conditions for which the motion vector prediction used for motion vector differential coding is zero. Justify all of them.

b) As you know, the solution recognized as most efficient for bitrate control with H.261 controls the quantization step as a function of the encoder output buffer fullness. Suggest a mathematical formula to express this dependency considering the values the quantization step may take for this recommendation. Indicate 3 necessary conditions that this formula must fulfill.

4.6) 1st Exam 1995/1996, 26 June 1996

Consider a videotelephone communications using Recommendation ITU-T H.261 for video coding at a rate of 64 kbit/s. The video sequence is coded with a CIF spatial resolution and a frame rate of 12.5 Hz.

The video content to code is horizontally divided in two equal parts but while the bottom part is fixed, the top part is moving. Since the encoder works processes sequentially the macroblocks, it is observed that all bits are uniformely generated in the first half of the time interval that the encoder usually dedicates to encoder each image. At the encoder, the bits wait for transmission in an output buffer. Knowing that the first image has used 15360 bits, the second image 20480 bit, and the third image 2560 bits, determine:

a) The instant of time at which the received has obtained all bits for the first image.

b) The minimum size of the encoder output buffer in order no bits are 'lost' in the situation described above.

c) The minimum visualization delay to apply at the decoder assuming that the encoder output buffer is the one determined in b) and the encoder may generate the bits with any distribution in the interval between the acquisition of two images.

4.7) 2nd Exam 2001/2002, 15 July 2002

a) Which is the global compression factor necessary to be able to transmit a video sequence with CIF spatial resolution (352×288 and 176×144 luminance and chrominance sample, respectively) at 10 Hz in a ISDN channel with 64 kbit/s, knowing that 10% of the overall available rate is used for synchronization and multiplexing data ?

b) Knowing that H.261 codes the quantization step with a fixed length code and supposing that the quantization step is no more send for all GOBs but only once per frame, which would be the bitrate saved for a 10 Hz CIF video sequence ? (Do not consider the fact that the quantization step may also be sent at macroblock level)

c) Indicate the two main reasons which justify that the most used rate control solution for H.261 is the variation of the quantization step.

4.8) 1st Exam 2005/2006, 20 June 2006

Consider the lab session about the Recommendation ITU-T H.261.

a) In the instants with higher video activity, notably at scene cuts, some macroblocks were classified in a particularly 'indesirable' way in terms of video decoded quality. What was this coding mode and why was it used ? What type of coding does it imply ? What distinguishes these macroblocks from others using the same H.261 coding mode without the same problems in terms of video quality and what is the subjective effect typical of these macroblocks ?

b) The existence of a encoder output buffer in H.261 codecs has, at least, two important impacts in terms of the quality of service provided to the final user; one impact is positive while the other is negative. What are these impacts and why do they happen ?

c) Assuming that an encoder is having difficulties to work in real-time, identify teo possible ways to address these difficulties while impacting as less as possible the final quality offered to the users.

5. Digital Video Storage, MPEG-1 and H.263

5.1) 2nd Exam 1993/1994, 8 July 1994

Consider the ISO/IEC MPEG-1 Video coding standard.

a) Do you think motion compensation may have the marginal effect of 'cleaning' the decoded image from artifacts due to the channel errors? If yes, give an example.

b) Explain for which reason this video coding standard includes three different types of coding at frame level or three frame types.

c) As you know, in this standard the various frame types use different coding tools and thus they should get different amounts of resources in terms of bitrate. Indicate, justifying, which criteria should be considered in terms of allocation of the available bitrate by the various frame types as a function of the amount of motion in the video sequences (at least two criteria).

d) Considering that the maximum number of bits spent per GOP is approximately the same as the average number of bits available per GOP e that only I and P frames are used, indicate which is the restriction to impose to the usage of I and P frames in order the maximum delay for random access does not exceed 250 ms (the CD reading speed is constant).

5.2) 1st Exam 1992/1993, 7 June 1993

Consider the ISO/IEC MPEG-1 Video coding standard.

a) In terms of coding tools, explain the fundamental differences between this algorithm and the one specified in Recommendation ITU-T H.261.

b) Explain which are the factors determining the selection of the N and M parameters for MPEG-1 video coding.

c) Explain the reason both the H.261 and MPEG-1 Video coding algorithms use relative macroblock addressing within the GOBs or slices (with the exception of the first macroblock transmitted).

5.3) 2nd Exam 1993/1994, 20 July 1994

Consider the MPEG-1 standard for digital video storage.

a) Indicate the reason why should the uniform distribution of bits by the various types of frames defined in this standard not be used.

b) Indicate, justifying, which is the main characteristic of the video decoder which dimensioning is strongly influenced by the implementation of the normal reverse mode? Explain why the same does not typically happen with the fast reverse mode.

c) Consider using the MPEG-1 Video coding algorithm at 25 Hz with M= 3 and N= 12 characterizing the temporal coding structure. If I frames get 3 times more bits than P frames and P frames get 4 times more bits than B frames (everything on average), compute the bitrate for the video data if, on average, each B frame macroblock uses 50 bits (each frame has 396 macroblock).

5.4) 1st Exam 1992/1993, 28 June 1993

Consider the MPEG-1 standard for digital video storage.

a) For which reason does the video coding algorithm specified in this standard uses three types of frames depending on the used coding tools?

b) Explain why the available bitrate is not equally distributed by the three types of frames.

c) Consider using the MPEG-1 Video coding algorithm with 1,5 Mbit/s at 25 Hz with M= 3 and N= 12 characterizing the temporal coding structure. If I frames get 3 times more bits than P frames and P frames get 4 times more bits than B frames (everything on average), compute the average number of bits available per macroblock for each type of frames assuming each frame has 396 macroblocks.

5.5) 2nd Exam1992/1993, 7 July 1993

The MPEG-1 Video standard basically uses the same coding tools used by Recommendation ITU-T H.261 and has a target bitrate falling within the bitrate range of H.261 (about 1.1 Mbit/s for video).

a) In this situation, which reasons motivated the specification of the MPEG-1 Video standard instead of just using H.261 already available ?

b) In terms of motion estimation and compensation, the MPEG-1 Video standard includes two big novelties regarding H.261. Which are these novelties ?

c) Consider using the MPEG-1 Video coding algorithm with 1,15 Mbit/s at 25 Hz with M= 4 and N= 16 characterizing the temporal coding structure. If I frames get 4 times more bits than P frames and P frames get 5 times more bits than B frames (everything on average), compute the average number of bits available per macroblock for each type of frames assuming each frame has 396 macroblocks.

5.6) 1st Exam 1994/1995, 18 June 1995

Consider the implementation of a digital TV system with the ITU-R 601 spatial and temporal resolutions (720×576 and 360×576 samples for luminance and chrominances, respectively, at 25 Hz).

a) Assuming that the transmission channel has a capacity of 100 Mbit/s and no compression algorithm is used, which is the maximum number of bits per sample that can be used to sample the luminance signal considering the luminance and chrominance sample use the same number of bits per sample.

b) Assuming now that a compression algorithm is used, providing compression factors of 20 and 25 for the luminance and chrominances, respectively, indicate which is the capacity of the transmission channel mentioned above which will remain free (consider ideal filters) for the number of bits per sample computed above.

5.7) 2nd Exam 1998/1999, 28 July 1999

Suppose you are contacted to project a high quality videoconference system for a multinational company. The maximum acceptable acquisition-visualization delay is 250 ms. The spatial resolution is 352×288 (Y) and 176×144 (Cr, Cb) at 25 Hz (8 bit/sample).

Suppose you have available, providing the necessary video quality two alternative systems: 1) a MPEG-1 Video codec with M=3 and N=3; 2) a H.261 codec. The average compression factors for the various frame types in the MPEG-1 Video codec are indicate in the table below, The average H.261 compression factors are the same as the MPEG-1 Video compression factors for the P frames. The first frame in H.261 coding as a compression factor similar to the MPEG-1 Video I frames and no P frame uses more bits than this frames

Frame Type	Luminance Compression Factor	Chrominance Compression Factor
I	12	20
P	15	25
B	25	35

Indicate, justifying, which solution would you propose to your client to satisfy his/her needs knowing that he intends to minimize the transmission costs.

5.8) 2nd Exam 1997/1998, 6 July 1998

Suppose you are contacted by a music clips production company which intends to start using digital systems for video processing and storage. The company tells you that it pretends a system with big flexibility in terms of edition – maximum access delay to each frame lower than 1 second – and that it pretends to store the highest possible number of clips with 4 minutes in a disc with 10

GBytes. The disc reading speed is 2 Mbit/s (constant). The clips have ITU-R 601 (European) resolution this means 720×576 (Y) and 360×576 (Cr, Cb) at 25 Hz.

Assuming you have available:

* *A JPEG codec providing compression factors of 10 and 15 (worst case) for the luminance and chrominances, respectively, with acceptable quality.*
* *A MPEG-1 Video codec using I, P and B frames com a periodic, regular and symmetric temporal coding structure, providing compression factors of 12 and 15 for the luminance and chrominances, respectively, for the I frames, and compression factors of 25 and 30 for the luminance and chrominances, respectively, for the P frames (both worst cases). B frames have compression factors three times higher than I frames but to obtain those compression factors it s necessary to use coding anchors always immediately adjacent in time to the B frames.*

Indicate, justifying, which solution would propose to your client to satisfy his/her needs and state how many complete clips will you be able to store in the available disc.

5.9) 1st Exam, 2001/2002, 26 June 2002

Consider the MPEG-1 Audio standard.

a) Indicate three important parameters for the definition of the cost of a terminal and for which the three layers of the MPEG-1 Audio standard offer a different trade-off.

b) What is the utility of the psychoacoustic model for the audio coding specified by the MPEG-1 Audio standard ?

c) Indicate the two main differences between the first/second layers and the third layer of the MPEG-1 Audio coding standard.

5.10) 2nd Exam 2002/2003, 11 July 2003

Consider the MPEG-1 Audio standard.

a) Explain what the thresholds of hearing and pain are. Which is their impact in audio coding ?

b) Explain what the masking effect in audio coding is. Explain the difference between temporal and frequency masking.

c) What is the basic idea behind audio perceptual coding ? Why are the characteristics of the auditory human system explored here and not the characteristics if the vocal tract like in speech coding ?

d) Indicate and explain what consist of the two main forms of audio coding in the frequency domain.

5.11) 2nd Exam 2001/2002, 15 July 2002

a) Indicate five factors which are important for the selection of the type of storage support for digital video storage.

b) Which are the two main temporal informations transmitted in a MPEG-1 Systems stream ? Explain for which frame type these two temporal informations are different ?

c) Which was the target quality for the MPEG-1 video storage systems ? Why ?

d) In which way and why are motion vectors coded in the MPEG-1 Video standard ?

e) Why does the MPEG-1 Video standard indicate at frame level the variation range of the motion vectors ?

f) Explain the reason which motivated the MPEG-1 Video standard to adopt different order for the acquisition, transmission and visualization of the video data.

5.12) 2nd Exam 2004/2005, 22 July 2005

Consider the MPEG-1 Audio standard.

a) Determine how many complete stereo songs with a 22 kHz bandwidth and 16 bit/sample with duration of 3 minutes is possible to store in a disk with 200 Mbytes if the coding is performed with MPEG-1 Audio Layer 3 to reach CD transparent quality.

b) Why does this standard use the DCT with overlapping window ?

c) What is the objective to use a scale factor for each audio subband ? What would happen if the coding would not use these scale factors ? What is the main difference between Layers 1 and 2 in terms of the coding of these scale factors ?

d) Which main factors would you take into account (at least 3) to select one of the MPEG-1 Audio Layers to code the audio for a certain application ?

5.13) 1st Exam 2002/2003, 20 June 2003

Consider a videotelephony communication according to recommendation ITU-T H.263 working at 40 kbit/s. The video data is coded at CIF spatial resolution and at 10 Hz temporal resolution. Each frame to transmit is horizontally divided in four equal parts; while the second and last parts are fixed, the first and third parts have motion. The encoder makes a sequential coding of the macroblocks and the bits are uniformly generated in the time periods where bits are produced; there are no bits generated in the time periods corresponding to the fixed areas with the exception of the first frame where the bits are uniformly generated for all the frames. After the first frame, the upper part has a lower activity than the third part which translates in a bit production which for each frame is always three times lower than the bit production for the third part. At the encoder, the bits wait to be transmitted in the encoder output buffer.

Knowing that the first frame used 15000 bit, the second frame 20000 bit and the third frame 8000 bit, compute, justifying:

a) The time for which the receiver gets all the bits corresponding to the second frame.

b) The minimum size for the encoder output buffer in order that no bits are lost in the situation described above.

c) The minimum initial visualization delay to apply at the decoder assuming that the encoder output buffer is half the size the one in the previous question and the encoder produces now instantaneously the bits for each frame at the time of its acquisition (infinitely fast coding).

5.14) 1st Exam 1999/2000, 30 June 2000

Consider the video coding recommendations ITU-T H.261 and H.263 (without the options).

a) Explain the difference between these two recommendations in terms of the motion vectors prediction. Which advantage has one mode over the other? Give an example of this benefit.

b) Explain the main difference between these two recommendations in terms of the motion estimation and compensation. Which are the main advantage and disadvantage of this difference?

c) Indicate an example of the H.263 increased efficiency in terms of the overhead data.

5.15) 2nd Exam 1998/1999, 5 July 1999

Consider recommendations ITU-T H.261 and H.263.

a) Which are the main functions of each of each hierarchical layer in the H.261 bitstream syntax?

b) Why don't these recommendations define the DCT quantization decision levels? (R: Because this is not needed for interoperability)

c) Indicate the main differences (2) between the H.261 and H.263 recommendations in terms of the motion vectors amplitude variation and motion vectors limitations, notably if H.263 optional modes are used.

d) Which are the main advantages and disadvantages (2) in using a finer granularity for the motion estimation in the context of recommendation H.263?

5.16) 2nd Exam 1997/1998, 6 July 1998

a) Indicate four of the main blocks in the architecture defined in recommendation H.324. For two of them, indicate which recommendations define their way of working.

b) Which are the rules governing the way H.263 encoders and decoder work in terms of spatial resolution?

c) Which are the main differences between recommendations H.261 and H.263 in the coding of the motion vectors (excluding the H.263 options)?

6. Digital Television

6.1) 2nd Exam 1997/1998, 24 July 1998

Suppose that your company has been contacted to design the video system feeding the giant screen at Sony Plaza. The digital transmission will be in high definition - 1920×1152 (Y) and 960×1152 (Cr, Cb) at 25 Hz (8 bit/sample). Assume that your have available providing the necessary quality for each frame, a MPEG-2 Video codec reaching the compression factors indicated in the table below. To guarantee adequate random access, at least one frame has to be coded in Intra mode every 300 ms. Finally, to reach the compression factors in the table, no more than three B frames should be introduced in a row and a P frame must always be used between two I frames. Assuming the intention is to minimize the bitrate to reduce the transmission costs, compute:

Frame Type	Luminance Compression Factor	Chrominance Compression Factor
I	10	15
P	15	20
B	20	30

a) The best M and N values characterizing the (regular) temporal coding structure of I, P and B frames to be adopted.

b) The average bitrate associated to the coding structured determined above.

c) The initial visualization delay at the receiver assuming the transmission is made at rate determined above, the N value is the same and M=N, and the critical compression factors for the I frames (for the 'more difficult' frames) are 10% lower than the average compression factors indicated in the table (they are the same for the other frame types). Assume also that the coding and decoding times are negligible.

6.2) 2nd Exam 1998/1999, 28 July 1999

a) Explain the main difference between the MPEG-1 and MPEG-2 standards in terms of the content they are able to code. What justifies this difference?

b) Explain which are the main reasons (2) motivating the specification of profiles and levels in the MPEG-2 Video standard.

c) Explain how it would be classified, in terms of profiles and levels, a decoder which capacity is below (even if very close) a certain profile@level conformance point. Why?

6.3) 2nd Exam 1997/1998, 6 July 1998

Consider the MPEG-2 standard for digital video coding.

a) Which were the main reasons (2) leading to the specification of profiles and levels in the context of this standard?

b) How and why is a bitstream classified in terms of profiles and levels if its characteristics are just slightly above a certain profile@level conformance point?

c) What would happen if a TV radio transmitter emits MPEG-2 coded video compliant to Main@Main and the receiver is compliant to the Simple@High? How does the situation change if the application context is now a multimedia database with online and offline MPEG-2 video coding?

6.4) 1st Exam 2001/2002, 26 June 2002

a) Which is the rationale delimiting the frontier between the technologies standardized by MPEG and DVB through ETSI? Why?

b) Indicate three important differences between the program and transport streams specified in the MPEG-2 Systems standard.

c) How many different Program Association Tables typically exist in a MPEG-2 transport stream? Why?

d) The MPEG-2 Video standard classifies each image to code as image-frame or image-field. Which is the difference between these two image types? Which is the motivation to define these two image types? In which conditions are these two image types typically used for interlaced content?

e) Which MPEG-2 Video conformance points did DVB adopted and with which objectives? What would happen if DVB had not adopted levels but just profiles in the context of MPEG-2 Video?

6.5) 2nd Exam 2001/2002, 15 July 2002

a) Explain what measures the so-called coding rate in systems like DVB-T.

b) Indicate which is the main idea used in the OFDM modulation adopted in DVB-T to reduce the number of mutually interfering modulated symbols.

c) What is the purpose of the guard interval used in the DVB-T modulation system? Which is conceptually the minimum value of this interval?

d) Indicate, justifying, which of the COFDM DVB-T variants is more indicated to cover small areas.

6.6) 2nd Exam 2003/2004, 19 July 2004

Consider the DVB standards.

a) Explain how the hierarchical 64-QAM modulation is used with benefit and for which conditions there are particular advantages in the adoption of this type of modulation.

b) Identify two relevant ways of combining the hierarchical 64-QAM modulation with the scalable coding methods available in MPEG-2 Video.

c) Identify two relevant factors for the selection of the encryption method to be used in the conditional access module of a digital pay TV system.

6.7) 2nd Exam 2005/2006, 8 July 2006

Consider a DVB-T digital TV system.

a) Determine how many complete stereo songs with a 22 kHz bandwidth and 16 bit/sample with duration of 3 minutes is possible to store in a disk with 200 Mbytes if the coding is performed with MPEG-1 Audio Layer 3 to reach CD transparent quality.

b) How many bytes do you have to spend for the coding of a 4:2:0 colour video sequence with 30 minutes and a 720×576 luminance spatial resolution at 8 bit/sample at 25 Hz if the luminance compression factor is 30 and the chrominances compression factors is twice that value ?

c) Knowing that a DVB-T solution may 'insert' 5 Mbit/s of source bits in a 8 MHz bandwidth channel, determine what would be the bitrate that may be 'inserted' if all the system parameters stay the same with the exception of the channel coding ratio that goes from ½ to 2/3 and the modulation that goes from QPSK to 64-QAM.

d) Which solution would you select from 1) 64-QAM modulation and ½ channel coding ratio and 2) QPSK and 5/6 channel codimg ratio, with all other parameters the same, if the selection criterion was the maximization of source bitrate. Knowing that the solution was to be used in a satellite transmission, would you remain more or less confident on the solution ? Why ?

7. Advanced Multimedia Coding

7.1) 2nd Exam 2005/2006, 8 July 2006

Consider the H.264/AVC video coding standard.

a) Regarding the slide groups defined in H.264/AVC, describe two ways to take benefit of this concept-

b) Explain the main benefit in defining type 2 for the frame coding mode when coding interlaced video content.

c) Explain why this standard uses an additional 4×4 Hadamard transform for the luminance DC coefficients from the macroblocks coded with the 16×16 Intra coding mode.

d) Identify the main difference, in terms of coding standards, between deblocking filters inside and outside the prediction loop.

7.2) 1st Exam 2005/2006, 20 June 2006

a) Identify three advantages in using an object-based coding model for visual representation such as the one defined in the MPEG-4 standard.

b) What is the main conceptual novelty adopted in the H.264/AVC coding standard for the coding of the Intra macroblocks ?

c) Explain what is the H.264/AVC Constrained Intra coding mode and why was defined.

d) Identify three reasons in terms of temporal prediction that may justify the substantial increase in the H.264/AVC encoding complexity.

7.3) 2nd Exam 2004/2005, 22 July 2005

a) DVB as decided to adopt the H.264/AVC video coding standard. Which have been the H.264/AVC profiles adopted and for which conditions. For what reason did DVB adopt two different H.264/AVC profiles ?

b) Identify three main factors which must have been considered by DVB when adopting the new H.264AVC profiles in addition to the previously adopted MPEG-2 Video profile.

c) Which are the main benefits (1) and drawbacks (1) of the H.264/AVC standard regarding the MPEG-2 Video standard ?

7.4) 1st Exam 2008/2009, 24 June 2009

Consider the scalable coding of video content.

a) Explain in your own words the type of functionality that it is possible to obtain with a scalable coded stream (and it is not possible to obtain with a non-scalable coded stream).

b) Indicate and justify what type of networks (at least 2) may take benefit from the usage of scalable coded video.

c) Explain what type of scalability does not normally imply any onus in terms of compression efficiency.

d) Indicate and justify the main advantage and drawback of using a scalable coding stream in comparison with an alternative simulcasting solution.

7.5) 2nd Exam 2008/2009, 20 July 2009

Consider that you become one of the YouTube managers in terms of multimedia content.

a) Indicate what disk capacity would you need to store 5.000.000 clips with an average duration of 4 minutes if the video has a spatial resolution of 352×288 luminance samples, 4:2:0 subsampling format, 25 Hz, the audio (stereo) a sampling frequency of 48 kHz (using the typical number of bit/sample), knowing that the audio compression factor is the typical MP3 compression factor for MP3 (for transparent quality) and the video compression factor is twice the audio compression factors for all video components.

b) How much would that capacity change (in percentage), if the spatial resolution would be reduced to half in both directions and the audio sampling frequency would become 40 kHz also knowing that in that case all compression factors would be reduced by 25 %.

c) For the situations above, explain what video coding standard would you use, idnicating 2 major system requirements justifying your choice.

d) If you were given the possibility to improve the overall user experience in a very significative way using 1% more of capacity, how would you invest this capacity and why ?

Solutions for the Exercises

1.1 None is right.

1.2 a) 24.06 MHz; b) 6.015 MHz.

1.3 a) Luminance (Y) and two chrominances (R-Y e B-Y); d) Alternate chrominances.

1.4 a) 10.6 MHz; b) Higher horizontal resolution.

1.6 a) Limited visual acuity (limited capacity to see details), capacity to obtain the illusion of motion, lower sensitivity to the chrominances regarding the luminance; b) Flicker; c) Simultaneous or alternative chrominances for the lines; no colour mixes.

1.7 c) 15 channels.

2.1 a) 15.8 and 0.78; c) In principle, yes.

2.2 a) Modulation without compression versus modulation with compression (vestigial band); b) Analogue versus digital; c) Analoguer versus digital; d) Telephone network versus data/computer network.

2.3 a) Vertital, horizontal, passage e without compression coding modes.

2.4 a) Control if any word is the start of any longer word; if yes, the conclusion would be that the codewords could not have been generated with Huffman entropy coding; b) Vertical, horizontal e passage and without compression; c) EOL words and MHM coded lines.

2.5 a) 2400 Hz; c) Because there are two code tables and synchronism has to be always kept.

2.6 a) k>9; b) 1.62 s; c) 311.04 bits.

2.7 a) 23.48; b) 16.67 and 28.32.

3.1 a) 48; b) 6; c) 82.

3.3 a) Sequential, progressive, hierarchical and lossless.

3.4 a) 6.048 s; b) 1.512 s, 4.84 s and 15.48 s.

3.5 a) 0.75 and 0.733.

3.6 a) Progressive and hierarchical; b) 32.48 s.

3.8 a) 3; b) 4; c) Quantization; d) Obtain finer granularity quality scalability.

4.1 a) 12441600 bit/s; c) 25.7.

4.4 a) 28800 bit e 550 ms; b) 6400 bit.

4.6 a) 240 ms; b) 28160 bit; c) 520 ms.

4.7 a) 211.2; b) 550 bit/s; c) Speed and granularity of reaction.

4.8 c) Using fast motion estimation methods and reducing the size of the motion window search.

5.1 a) Yes; c) At least four I frames per second if they are evenly distributed in time.

5.2 a) Bidirectional motion compensation and half-pixel motion vectors accuracy.

5.3 a) Because the various types of frame have different compression capabilities what would prevent obtaining constant quality; b) Memory; c) 1.32 Mbit/s.

5.4 b) Due to the different compression capabilities of the three types of frames and the need to obtain (almost) constant quality; c) NB = 56.8; NP = 227.3; NI = 681.8.

5.5 a) To addresses new requirements; b) Bidirectional motion compensation and half-pel motion vectors accuracy; c) NB = 39.54; NP = 197.72; NI = 790.89.

5.6 a) 4 bit/sample; b) 94.4 Mbit/s.

5.7 MPEG-1 Video.

5.8 MPEG-1 Video.

5.9 a) Quality, complexity and bitrate; b) MDCT transform and Huffman entropy coding.

5.11 a) Capacity, reading speed, access time, type of access, durability, mobility, cost; b) PTS and DTS.

5.12 a) 75; b) Reduction of the block effect; d) Required quality; acceptable complexity; acceptable delay; available bitrate.

5.13 a) 875 ms; b) 32000 bit; c) 400 ms.

5.14 b) Half-pel motion estimation and compensation; c) GOB headers don't have to be sent for 'empty' GOBs.

5.15 b) Not needed to guarantee interoperability.

5.16 a) Video codec with recommendations ITU-T H.261 or H.263 and Multiplex with recommendation ITU-T H.223.

6.1 a) N=6, M=3; b) 45.47 Mbit/s; c) 305 ms.

6.2 a) Progressive versus interlaced content; b) To provide interoperability with limited complexity; c) Profile and level immediately below.

6.3 a) Providing interoperability and limiting complexity; b) Profile and level immediately above; c) Decoder is not able to decode since it does not know some of the coding tools eventually used.

6.4 c) 1; e) Main@Main e Main@High.

6.5 a) Fraction of the total rate used for source data; d) 2k.

6.7 a) 75; b) 777.6 Mbytes; c) 20 Mbit/s; d) Solution 1.

7.3 a) Main for SDTV and High for HDTV; b) Increase in the compression capacity; increase in complexity; compatibility with existing terminals; c) Better compression efficiency and higher complexity.

7.4 b) Mobile networks and the Internet; c) Temporal scalability.

7.5 a) 1.67×1015 bits; b) -59.4 %; c) H.264/AVC; d) Metadata.