MULTIMEDIA COMMUNICATION

MEEC, MECD, METI & MEE

EXERCISES

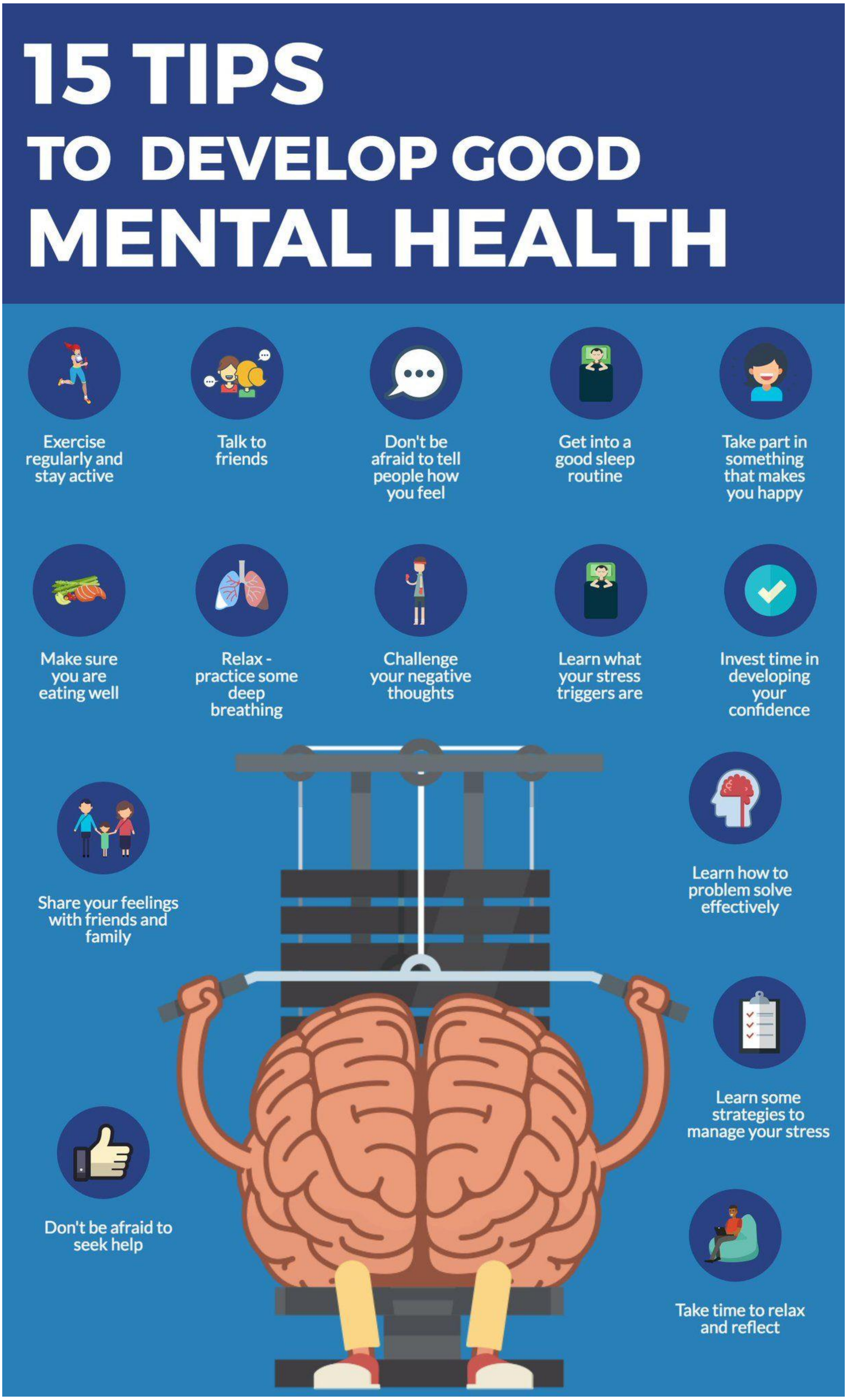
(with abbreviated solutions)

Fernando Pereira

INSTITUTO SUPERIOR TÉCNICO

Departamento de Engenharia Electrotécnica e de Computadores

September 2023


15 TIPS
TO DEVELOP GOOD
MENTAL HEALTH
Exercise regularly and stay active
Talk to friends
Don't be afraid to tell people how you feel
Get into a good sleep routine
Take part in something that makes you happy
Make sure you are eating well
Relax - practice some deep breathing
Challenge your negative thoughts
Learn what your stress triggers are
Invest time in developing your confidence
Share your feelings with friends and family
Learn how to problem solve effectively
Learn some strategies to manage your stress
Don't be afraid to seek help
Take time to relax and reflect

1. Basics

1.1) 23/06/2021

a) Explain why any analogue multimedia signal becomes 'richer' when its bandwidth increases.

b) Explain why any digital multimedia signal becomes more 'complex' when the corresponding analogue multimedia signal increases its bandwidth.

c) Explain what mainly determines the number of bits per sample when digitizing an analogue multimedia signal, e.g. speech, audio, video. Why is this number different for speech and audio ?

d) For video, what is the key factor determining a proper frame rate ? What happens to the user experience if the frame rate is too low ?

e) For video, how should the spatial resolution change if the display size increases OR the viewing distance reduces ?

1.2) 17/06/2019

a) How many signal components are needed to see a gray image in a black and white television ? Which ones ?

b) How many signal components are needed to see a gray image in a colour television ? Which ones ?

c) How many signal components are needed to see a colour image in a colour television ? Which ones ?

d) In total, are there more samples or pixels in a black and white image ? And in a colour image ? Why ?

e) Are there more luminance samples or total chrominance samples in a 4:2:2 colour image ? Why ?

1.3) 28/01/2020

a) Indicate two fundamental differences between the luminance signal of an analogue and a digital image.

b) Why do humans see colors less well during the night ?

c) What is a masking effect in images ?

d) What fundamentally defines the number of frames per second needed for a video signal ?

e) How does the number of pixels in an image vary when the chrominance subsampling factor changes from 4:4:4 to 4:2:2 and 4:2:0 ?

f) Assuming you have to show images under multiple different conditions, e.g. display distance to viewer and display size, would you prefer images with increased spatial resolution or increased number of bits per sample ? Why ?

1.4) 23/07/2019

a) What is the difference between the bandwidth and the sampling frequency of a video signal ?

b) Indicate two benefits for Humans to have binocular vision.

c) What property relates the R, G and B values of a grey pixel ? Why ?

d) What are the chrominance (U and V) values of a grey pixel ? Why ?

e) Why do video services typically transmit chrominances (UV) with luminance and not directly primary colors (RGB) ?

f) Why do video services typically use 4:2:0 chroma subsampling ?

2. Imaging

2.1) 20/07/1994

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 2 Mbit/s transmission channel.

a) Considering that the transmission channel is available during 10 s, how many complete bi-level images can be transmited without any compression ?

b) And how many complete grey images can be transmited, in the same 10 s, if images with 128 grey levels are used (still 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 be transmited still in the same 10 s ?

2.2) 08/07/1994

Consider the JPEG compression standard for digital images.

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

b) Considering now that the images are coded with the hierarchical mode, determine 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.*
- *always 8 bit/samples.*
- *the compression factors for each layer increase 25 % for each new layer regarding the previous layer.*

- *the compression factors for the base layer are those indicated in a).*

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

2.3) 08/07/2006

a) Determine the average number of bits per pixel used (considering the luminance and chrominances) 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 is used for the chrominances.

b) What is the main difference between a lossless and a lossy image coding system ? Which of these 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 largest disadvantage of entropy coding for transmissions in mobile environments ?

2.4) 20/06/2006

Assume that a user wants to access a database with JPEG coded images to search for some specific 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:

   1. for sequential coding, the average compression factors for the luminance and chrominances are 10 and 15, respectively
   2. for the base layer of the progressive coding mode, the compression factors are twice as high the factors for sequential coding
   3. 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 a target image in the maximum quality if images with 2 layers are used and all image layers are sequentially transmited unless the user stops the transmission by browsing to 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.

2.5) 29/06/2012

Consider the JPEG standard to code photographic images.

a) Determine the compression factors that would be needed for the luminance and for the chrominances to spend an average number of 0.64 bit/pixel (considering both the luminance and the chrominances) when coding a 4:2:0 image with 8 bit/sample, knowing that the average luminance compression factor is twice the average chrominances compression factor.

b) Determine the total number of bits that have to be spent to code a 720×576, 4:2:2, 8 bit/sample image if an average number of 3 DCT coefficients are coded per block and each coefficient costs, on average, 4 bits; additionally consider that the EOB (End of Block) word costs 2 bits and all blocks in the image spend bits.

c) Consider a 4:2:2, 8 bit/sample image coded with the hierarchical mode. How many layers can we use to code the image if the base layer is 720×576 for the luminance and has a global (luminance and chrominances) compression factor of 20, the global compression factor doubles for each new layer, each new layer has twice the resolution in both directions, and the total number of bits spent should be less than 10^6 bits.

2.6) 28/06/2004

Consider the JPEG standard to code digital images.

a) Identify and explain the working process of the two JPEG ways of implementing the progressive coding mode.

b) How would you select the prediction mode to use in the context of the JPEG lossless coding mode ? What would you do if there was a need to guarantee that the effect of transmission errors do not propagate too much in the decoded image ?

c) Which are the main benefits and drawbacks of using the VLI codes to code the amplitude of the DCT coefficients in the JPEG baseline coding process ?

d) Explain the relevance of the concept of entropy of a source for the designer of a source encoder.

2.7) 09/07/2019

Consider the coding of digital images.

a) Assuming the most typical number of bits per sample, what would be the compressed rate cost of one luminance sample and of one chrominance sample (separately) if the compression factors for the luminance and chrominance are 20 and 40, respectively ?

b) How many bits would cost, on average, a JPEG luminance block for the situation described in a) ?

c) How many bits would cost, on average, a JPEG luminance image with resolution 576×720 for the situation described in a) ?

d) Assuming that 280000 bits are available to code a 4:4:4 image for the situation in a), what would be its maximum horizontal resolution knowing that its vertical resolution is 500 lines ?

e) Assuming that the same 280000 bits are available to code a 4:2:0 image, and the same situation as in a) applies, how would the total number of pixels in the image change (same, increase or decrease) in comparison with the situation in d) ?

2.8) 23/06/2021

Consider the JPEG standard to code photographic images with a 576×720 luminance resolution, 4:2:0 color subsampling and 8 bit/sample.

a) How many times would the number of luminance blocks increase if the color subsampling is changed to 4:2:2 ? And the total number of chrominance blocks for the same change ?

b) Determine the average number of bits per pixel that has to be spent to code this type of image if a codec with a luminance compression factor of 25 and a chrominances compression factor of 15 is used.

c) Determine the chrominances compression factor if it is required that the rate cost for the luminance of one pixel is the same as for the two chrominance components of one pixel if the luminance compression factor is 25, assuming the color subsampling is now 4:4:4.

d) Determine the maximum number of DCT coefficients that may be coded for each luminance block of an image if each coefficient costs, on average, 4 bits for the luminance and 3 bits for the chrominance and a maximum total spending of 350000 bits for one image is desired; consider that luminance blocks always code 2 coefficients more that each chrominance block and additionally consider that the EOB (End of Block) word costs 3 bits.

3. Video

3.1) 22/06/1994

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

a) State for which reason motion estimation and compensation are 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 standardized 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 ?

3.2) 08/07/1994

Consider the video compression algorithm for videotelephony and videoconference specified in Recommendation ITU-T H.261. For 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

3.3) 20/07/1994

Consider a videotelephony communication using Recommendation ITU-T H.261 at 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 sequence 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 (naturally the bits in a) are not relevant anymore here).

3.4) 30/06/1995

Consider Recommendation ITU-T 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, a rather efficient bitrate control solution for 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 that the quantization step may take for this coded. Indicate 3 necessary conditions that this formula must fulfill.

3.5) 26/06/1996

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.

Each video image to code is horizontally divided into two equal parts but while the bottom part is fixed, the top part is moving. Since the encoder 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 encode 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 bits, and the third image 2560 bits, determine:

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

b) The minimum size of the encoder output buffer in order the bits mentioned above are appropriately transmitted.

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.

3.6) 15/07/2002

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 justifying that the most used rate control solution for H.261 is the variation of the quantization step.

3.7) 20/06/2006

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

a) In the time 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 typically associated to these macroblocks ?

b) The existence of an 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 two possible ways to address these difficulties while impacting as less as possible the final quality offered to the users.

3.8) 11/06/2012

Consider a videotelephony communication using Recommendation ITU-T H.261. The video sequence is coded with a CIF spatial resolution and a frame rate of 12.5 Hz at a rate of 128 kbit/s. The video content to code is horizontally divided into two equal parts; however, while the bottom part is fixed, the top part is moving. Since the encoder processes sequentially the macroblocks, it is observed that all bits are uniformly generated in the first half of the time interval that the encoder usually dedicates to encode 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 10240 bits, determine:

a) The time instants at which the receiver obtains all bits for the first, second and third images.

b) The minimum size of the encoder output buffer in order all bits above are transmitted without problems.

c) The initial visualization delay associated to the system defined in b).

d) The maximum number of bits that the 4th image may spent (still assuming that it only spends bits in the top half).

3.9) 24/06/2009

Consider a ITU-T H.261 videophone system coding video with a spatial resolution of 352×288 pixels for the luminance, 4:2:0 chrominance subsampling, with 8 bit/sample), at 12.5 Hz. Assume that the average compression factor (*measured over all the macroblocks in the image*), without any external constraints in terms of coding modes, is 25 for the luminance and 30 for the chrominances (header bits not included). Assume that 500 bits of overhead are spent per frame.

a) Assuming that for each frame, on average, only 200 macroblocks generate code bits (the remaining ones are so similar to the previous image that no update is needed), determine the average overall compression factor (including luminance and chrominances) measured over the macroblocks which effectively generate code bits, including also the overhead bits.

b) If for editing reasons, all the macroblocks of all frames had to be coded in intra mode, what would be the total bitrate assuming that the compression factor for this type of coding is half the compression factors indicated for the luminance and chrominance ?

c) For the situation in a), assume that to guarantee a higher error protection, one out of each 100 macroblocks spending bits is necessarily coded in intra mode. Assuming that the intra coding mode has a compression factor as defined in b), determine what would be the global compression factor (considering all the macroblocks in the image) corresponding to this situation also including the header bits.

d) If a transmission rate of 1 Mbit/s is used, what would be maximum number of bits that the first frame may spend if a maximum acquisition-visualization delay of 200 ms would be requested.

Assume that the encoder generates the bits for each frame uniformly in the time period between the acquisition moments of each two successive frames.

3.10) 09/07/2019

Consider a videotelephony communication using Recommendation ITU-T H.261. The video sequence is coded with a CIF spatial resolution (352×288 samples for the luminance, 4:2:0), a frame rate of 10 Hz and a constant bitrate channel of 64 kbit/s. The output buffer has a size of 12800 bits. The bits for each coded image are uniformly generated in the time between the acquisition of two images. Answer the following questions, always explaining.

a) What is the average number of bits that each image captured by the camera may spend ?

b) What is the maximum number of bits that each image captured by the camera may spend ?

c) What is the minimum number of bits that each image captured by the camera may spend ?

d) What is the maximum number of bits that the third image may spend ?

e) What is the maximum number of bits that a image may spend if the buffer is full when the image starts to be coded ?

f) What is the maximum number of motion vectors that may be used to code a single image ?

g) Considering that a constant bitrate channel is used, what architectural element takes the most important decisions to allow the encoder controlling the number of bits spent per frame ? Why?

3.11) 20/07/1994

Consider the MPEG-1 standard for digital video storage.

a) Indicate the reason why it is not advisable to uniformly distribute the available bits by the various types of frames defined in this standard.

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 (always on average), determine the bitrate for the video data if, on average, each B frame macroblock uses 50 bits (each frame has 396 macroblocks).

3.12) 07/07/1993

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 simply using the H.261 solution 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 (always on average), compute the average number of bits available per macroblock for each type of frames assuming that each frame has 396 macroblocks.

3.13) 28/07/1999

Suppose that 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).

Assume that 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 indicated 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 has a compression factor similar to the MPEG-1 Video I frames and no P frame uses more bits than these frames.

Frame Type	**Luminance CompressionFactor**	**Chrominance CompressionFactor**
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/she intends to minimize the transmission costs.

3.14) 06/07/1998

Suppose that 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 large flexibility in terms of edition – maximum access time for 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 that you have available:

1. *A JPEG codec providing compression factors of 10 and 15 (worst case) for the luminance and chrominances, respectively, with acceptable quality.*
2. *A MPEG-1 Video codec using I, P and B frames with 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 (always worst cases). B frames have compression factors three times higher than I frames but to obtain those compression factors it is 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 would you be able to store in the available disc.

3.15) 12/07/2010

Consider using the MPEG-1 Video codec to code CIF (396 macroblocks) video information at 25 Hz to be stored in a CD. Assume that M=3 is used and the I get 3 times more bits than the P frames while the P frames get 4 times more bits than the B frames (always on average). The average number of bits per macroblock in a B frame is 50.

For the conditions above, determine the acceptable set of N values if it is requested that the video bitrate does not exceed 1.8 Mbit/s and the maximum access time does not exceed 400 ms. Assume that the reading rate is the same as the coding rate.

3.16) 29/07/2012

Suppose that you have been contacted by a company to design a videoconference solution to work between the various EURO'2012 stadiums using the lowest possible bitrate while guaranteeing the necessary minimum video quality. The company also requires that the initial visualization delay (measured as the maximum time difference between corresponding acquisition and visualization instants) is below 300 ms. The video resolution is CIF (352×288 luminance samples), 4:2:0 at 12.5 Hz with 8 bit/sample. Assume that you have available and providing the necessary minimum video quality two systems:

1. H.261 system with average compression factors of 18 and 22 for the luminance and chrominances, respectively, and critical compression factors (for the most difficult images) of 12 and 15 for the luminance and chrominances, respectively.
2. MPEG-1 Video system with N = M = 3 and average compression factors of
    a. 18 and 22 for the luminance and chrominances, respectively, in the I frames
    b. 25 and 35 for the luminance and chrominances, respectively, in the B and P frames

The critical compression factors are 75 % of the average compression factors for all frame types.

a) Determine which of the solutions above you would select to better satisfy the needs of your client assuming that the transmission rate for each solution correspond to its coding rate.

b) Determine for which transmission bitrates the MPEG-1 solution would satisfy the initial visualization delay requirement, assuming that for this solution the transmission rate may be regulated (meaning that the coding and transmission rate would be different).

c) Explain 2 performance impacts of increasing the value of M, this means increasing the number of B frames between two anchor frames.

3.17) 24/07/1998

Suppose that your company has been contacted to design the video system feeding the giant screen at Sony Plaza at EXPO'98. The digital transmission will be in high definition - 1920×1152 (Y) and 960×1152 (Cr, Cb) at 25 Hz (8 bit/sample). Assume that you 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 consecutively and a P frame must always be present between two I frames. Assuming that the intention is to minimize the bitrate to reduce the transmission costs, determine:

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 that the transmission is made at the 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.

3.18) 28/07/1999

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?

3.19) 06/07/1998

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 Simple@High? How does the situation change if the application context is now a multimedia database with online and offline MPEG-2 video coding?

3.20) 26/06/2002

a) What 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. What is the difference between these two image types? What was the motivation to define these two image types? In which conditions are these two image types typically used for interlaced content?

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

3.21) 08/07/2019

Suppose that you are contacted by an advertising company to design a multimedia digital storage system. The maximum access speed to the disk is 30 Mbit/s. The clips have HDTV resolution, this means 1920×1152 (Y) and 960×1152 (Cr, Cb) at 25 Hz. Assume that you have at your disposal providing the required video quality:

1. a JPEG coding solution with a compression factor of 25 for both the luminance and chrominances

2. a MPEG-2 Video coding solution with the following compression factors:

· I frames: 30 and 35 for the luminance and chrominances, respectively

· P frames: 40 and 50 for the luminance and chrominances, respectively

· B frames: 50 and 60 for the luminance and chrominances, respectively

a) If your client asks for the most compression efficient frame level random access solution (meaning that each frame should be independently accessed), what solution would you offer him/her from those above ?

b) If your client asks for the most compression efficient, very low complexity coding solution, what solution would you offer him/her from those above ?

c) If your client asks for a coding solution with best interoperability with the current image-coding ecosystem, what solution would you offer him/her from those above ?

d) If your client asks for the most compression efficient coding solution with a maximum access time per image below 400 ms, what solution would you offer him/her from those above if M=2 had to be used for the MPEG-2 Video solutions (if relevant, specify the GOP size) ?

e) If you could have available a third coding solution, notably to increase the compression efficiency, what solution would you like to have at your disposal ? Why ?

3.22) 08/07/2006

Consider the H.264/AVC video coding standard.

a) Regarding the slice groups defined in H.264/AVC, describe two ways of taking benefit of this concept.

b) 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.

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

3.23) 20/06/2006

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

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

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

3.24) 24/06/2009

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.

3.25) 20/07/2009

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, indicating 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 ?

3.26) 11/01/2019

Consider the video codec specified in H.264/AVC standard.

a) What may be service implications of the fact that H.264/AVC spends “about 50% less rate for the same perceptual quality regarding previous existing standards” (2 implications) ?

b) Why is it appropriate to say that H.264/AVC “does NOT allow to guarantee any minimum level of quality” ?

c) How does H.264/AVC try to overcome the limitations of using a translational motion model ?

d) What is the main goal of using well selected half- and quarter-sample interpolation filters ?

e) What are the main positive and negative impacts of using multiple reference frames ?

f) What is a main practical impact (different from the past) of including in List 0 and List 1 the same (already decoded) frames although in a different order ?

3.27) 08/07/2021

Consider the H.264/AVC video coding standard.

a) What is the offered compression factor if this codec spends around 2Mbit/s to code video with a spatial resolution of 576×720, 4:2:0 at 25 Hz ?

b) Explain what is a practical implication of using the Flexible Macroblock Ordering tool.

c) Why are motion vectors differentially coded but not across slices ?

d) Explain why Instantaneous Decoding Refresh (IDR) pictures and not just Intra pictures are essential to allow zapping in a broadcasting scenario.

e) For what reason was the Constrained Intra Coding Mode defined ?

f) Explain why is the H.264/AVC transform process called a hierarchical transform ?

4. Audio

4.1) 26/06/2002

Consider the MPEG-1 Audio standard.

a) Indicate three important parameters for the definition of the cost of a terminal 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 solutions specified by the MPEG-1 Audio standard ?

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

4.2) 11/07/2003

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

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

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

4.3) 22/07/2005

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 process 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 ?

4.4) 29/06/2012

Consider the MPEG-1 Audio standard to code audio content with 22 kHz bandwidth; assume reasonable compression factors and the most usual number of bits per sample.

a) How many complete stereo music pieces, with a duration of 4 minutes, can we store in a 900 MBytes disk using the Layer 3 of the MPEG-1 Audio standard to code the music content with a transparent quality regarding CD music content.

b) What is the maximum duration of each music piece that we can afford if we want to store 1000 musics in the same disk as above using a Layer 2 MPEG-1 Audio codec?

c) Explain how would the maximum number of stored musics vary if we increase the audio bandwidth three times but the audio becomes mono and not anymore stereo.

d) Describe two main technical differences between the MPEG-1 Audio Layer 2 and Layer 3 codecs and the corresponding advantages.

4.5) 08/07/2019

Consider the MPEG-1 and MPEG-2 Audio standards.

a) Determine the coding rate for stereo audio content with a 22 kHz bandwidth and the usual number of bit/sample if coded with a Layer 3 codec to reach CD transparent quality. How would the rate vary in percentage if the sampling rate becomes 48 kHz and mono audio is used.

b) What does it mean saying that audio does not have a universal source production model ? Why is this different for speech ?

c) Why does the Layer 3 codec use the MDCT with an overlapping window ? How is this overlapping applied ?

d) Why does the Layer 3 codec use the MDCT with a varying size window ? How is this varying size window applied ?

4.6)

Consider audio signals.

a) Indicate two audio characteristics which are numerically larger than the corresponding speech characteristics.

b) What is the main audio characteristic justifying to use a higher bit depth (i.e. number of bits per sample) than for speech ? Explain the process involved.

c) Explain in what consists the temporal post-masking effect.

d) What is the main motivation to increase the size of the audio frames in MPEG-1 Audio Layer 2 regarding Layer 1? Which performance metric suffers if we keep increasing the audio frames size ?

e) Would you prefer listen to 22 kHz bandwidth with appropriate sampling and 16 bit/sample audio OR 44 kHz sampling frequency and 12 bit/sample audio ? Why ?

f) Considering the varied composition of a symphonic orchestra, how would change the subjective quality assessment associated to the music if the full bandwidth (22 kHz) is kept while successively reducing the sampling rate from a value which starts being 2.5 times the full bandwidth and ends being 1.5 times the full bandwidth ?

4.7) 18/01/2021

Consider the MP3 audio coding standard.

a) What is the first key consequence in terms of user experience if longer audio frames are coded with MP3 ? What type of positive impact could this have ?

b) What is the major impact in terms of coding parameters when audio masking increases the hearing threshold at a specific bandwidth range ? What would be an important practical impact at service level ?

c) Is lossy perceptual audio coding compatible with 'transparent quality coding' ? How ?

d) What would happen if a MPEG-1 Audio Layer 1 decoder is asked to decode a MP3 stream ? And a MPEG-1 Audio Layer 2 decoder?

4.8) 16/07/2020

Consider the audio codec specified in MPEG-1 Audio Layer 3.

a) What is a reasonable coding rate to code a stereo audio signal where each audio channel is sampled at 44 kHz if the codec above is used?

b) How would be a good estimation of the new coding rate above change if the audio bandwidth is halved ?

c) Indicate two coding effects that may happen if the audio encoder includes a psychoacoustic model which is ´conservative' in the sense of using lower than appropriate hearing thresholds associated to the audio frequency masking effects ?

d) Indicate two coding effects that may happen if the audio encoder includes a psychoacoustic model which is 'optimistic' in the sense of using higher than appropriate hearing thresholds associated to the audio frequency masking effects ?

e) Considering the varied composition of a symphonic orchestra, how would change the user audio subjective experience if the audio bandwidth is successively reduced/filtered while not reducing the initial sampling rate which respects the sampling theorem ?

5. 3D Video

5.1) 11/06/2012

As you know, 3D video is nowadays very popular.

a) Identify and explain the two main ways of providing the user a 3D video experience.

b) Define both and explain the difference between stereo and movement parallaxes. Which of these types of parallax may be present in a multiview video system with many views ? Why ?

c) Explain what does a requirement on view-switching random access typically ask for. When is this type of requirement important for an user ?

d) Compute the typical bitrate for a video stereo pair when using the Multiview Video Coding (MVC) with standard resolution if the two views are coded with similar PSNR. Compute the same bitrate for a system with 10 views.

5.2) 29/06/2012

Consider a 3D video system.

a) Explain what is a frame compatible stereo format. Also explain the difference between a spatial multiplexing and a time multiplexing frame compatible stereo format.

b) Explain why frame-compatible stereo video tends to have higher spatial frequency content characteristics.

c) What is the most important new feature/tool of the Multiview Video Coding (MVC) standard regarding the H.264/AVC standard ? How does it work ?

d) What is the implication of the 'backward compatibility' requirement for the MVC standard ?

e) If the backward compatible view in a MVC stereo pair spends 2 Mbit/s, what is the minimum rate that the second view has to spend if a perceptual quality similar to stereo simulcasting with 2+2 Mbit/s had to be achieved ? Why?

5.3) 28/01/2020

Consider a virtual reality system based on 360 degrees video.

a) What is the main difference between virtual reality and augmented reality experiences ?

b) What is the main purpose of performing stitching in this context ?

c) Why is the spherical visual data projected into a rectangle as when using the equirectangular projection ?

d) What is the so-called viewport ?

e) What is precisely the so-called motion to photon delay ?

f) What is the main visualization 'parameter' that the user can (indirectly) control when navigating the visual data, which is critical to select the visual data to be displayed ?

5.4) 29/01/2019

a) What is the retinal disparity ?

b) Why doesn't the accommodation-vergence conflict happen in real life ?

c) What would you do to the baseline of a stereo camera if you wanted to intensify the final depth effect ?

d) Explain why are occlusions a good hint/cue for 3D/depth perception? Why is it a monocular cue ?

e) What is the main advantage of exploiting interview redundancy as in conventional stereo coding ?

f) Why is it problematic to have large coding errors at the sharp edges corresponding to object borders in a depth map ?

Solutions and Just Ideas for Solutions

(in the exam you must explain and motivate your answer; a number or a word is NOT enough)

1. Basics

1.1 a) idea: more frequencies and faster variations; b) higher sampling rate, memory, etc.; c) idea: transparency regarding analogue version; different dynamic range; d) idea: illusion of motion; hiccup effect; e) must increase.

1.2 a) 1, luminance; b) 3, RGB or YUV; c) 3, RGB or YUV); d) the same; more samples; e) the same.

1.3 a) idea: digital signal only exists at periodic time instants and taking allowed quantization levels; b) idea: during night Humans see with rods since they are more sensitive than cone cells (100 times more) and they are sensitive to luminance not any specific color range; c) idea: the effect where one image stimulus hides other imade stimulus; d) the achievement of illusion of motion; e) does not change as only depends on the luminance resolution; f) increased resolution.

1.4 a) idea: the bandwidth corresponds to the maximum frequency present in the analogue signal while the sampling frequency corresponds to the rate samples are taken to convert it to a digital signal and is is defined by the Sampling Theorem as at least twice the bandwidth; b) idea: wider field of view and perception of depth; c) R,G,B values are the same; d) zero; e) idea: to save rate by exploiting the lower human sensitivity to chrominance; f) to reduce the number of samples to code by exploiting the lower human sensitivity to chrominance.

2. Imaging

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

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

2.3 a) 0.75 bit/pixel and 0.733 bit/pixel.

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

2.5 a) 25; 12.5; b) 181440; c) 2.

2.6 a) spectral selection, successive approximations

2.7 a) 0.4 and 0.2 bit/sample; b) 25.6 bit/block; c) 165888; d) 700 sample/line; e) increase.

2.8 a) does not change; doubles; b) 0.587 bit/px; c) 50; d) 9.

3. Video

3.1 b) to exploit the redundancy between neighboring motion vectors; d) idea: human visual system is very sensitive to the error in the DC component.

3.3 a) 28800 bit and 550 ms; b) 6400 bit.

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

3.6 a) 211.2; b) 550 bit/s; c) speed and granularity of the reaction.

3.7 a) Skip mode; c) using fast motion estimation methods and reducing the size of the motion window search.

3.8 a) 120, 280 and 360 ms; b) 20480 bits; c) 240 ms; d) 10240 bits.

3.9 a) 13.23; b) 1.145 Mbit/s; c) 25.94; d) 200 000 bits

3.10 a) 6400; b) 19200; c) 0; d) 19200; e) 6400; f) 396; g) Quantizer.

3.11 a) because the various types of frame have different compression capabilities and thus obtaining rather constant quality implies a non-uniform bitrate allocation; b) memory; c) 1.32 Mbit/s.

3.12 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.

3.13 MPEG-1 Video.

3.14 MPEG-1 Video.

3.15 N=6

3.16 a) H.261; b) > 1.024 Mbit/s.

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

3.18 a) progressive versus interlaced content; b) to provide interoperability with limited complexity; c) profile and level immediately below.

3.19 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.

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

3.21 a) MPEG-2 Video only with I frames; b) MPEG-2 Video only with I frames; c) JPEG; d) MPEG-2 Video with N=24; e) H.264/AVC.

3.24 b) mobile networks and the Internet; c) temporal scalability.

3.25 a) 1.67×10^{15} bits; b) -59.4 %; c) H.264/AVC; d) metadata.

3.26 b) since encoders are not normative, no minimum quality may be guaranteed; c) idea: variable and dynamic block size motion compensation; d) idea: to make better temporal predictions to reduce the prediction error and thus the rate for a target quality.

3.27 a) 62.2.

4. Audio

4.1 a) quality, complexity and bitrate; c) MDCT transform and Huffman entropy coding.

4.2 a) idea: the impact in coding is to define signals which do not need to be coded or should be 'limited' before coding; b) idea: auditory masking is the hearing behavior when the perception of one sound is affected (masked) by the presence of another sound; c) idea: exploit the characteristics of the human auditory syste to reduce the coding rate for a target quality; d) subband coding and transform coding.

4.3 a) 75; b) reduction of the block effect; d) required quality; acceptable complexity; acceptable delay; available bitrate.

4.4 a) 255; b) 40.91 s; c) reduction to 2/3.

4.5 a) 117.333 kbit/s; -45%.

4.6 a) bandwidth and dynamic range; b) larger dynamic range; e) first solution; f) reducing quality due to aliasing

4.7 a) idea: more delay; less rate as more temporal redundancy to exploit; b) idea: quantization step increase; less rate for same quality; c) yes, by removing only perceptually irrelevant information; d) idea: No decoding would be possible as the later codecs have fewer coding tools and thus a simpler coding syntax.

4.8 a) 177.333 kbit/s; b) 58.666 kbit/s; e) quality is reduced.

5. 3D Video

5.1 d) 3.4 Mbit/s and 15.5 Mbit/s.

5.2 e) 500 kbit/s.

5.3 e) user viewing direction.

5.4 c) increase the camera baseline; e) reduce the coding rate.