# MULTIMEDIA COMMUNICATION

## INSTITUTO SUPERIOR TÉCNICO

Academic Year 2019/2020 – 2nd Semester, Responsible: Prof. Fernando Pereira

**2nd Exam – 16th July 2020 (Thursday), 11.30am**

The marks should be out before **17th July (Friday), 8pm** at the CMul Web page. If you want to check the exam scoring, please contact the course responsible by email to schedule a conversation.

The exam is **2.5 hours long**. Answer all the questions in a detailed way, including all the computations performed and justifying well your answers.

*Don't get 'trapped' by any question; move forward to another question and return later.* ***Good luck !***

I (1.0 + 1.0 + 0.5 + 0.5 + 1.0 val. = 4.0 val.)

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

a) How many **pixels, samples and blocks** exist in this type of image. (R: 414720, 1244160, 19440)

b) Determine the **price in bits per pixel** if a codec with a luminance compression factor of 20 and a chrominances compression factor of 10 is used and 10 bit/sample is used. (R: 2.5 bpp)

c) Determine the **price in bits for a full image** if a codec with a luminance compression factor of 25 and a chrominances compression factor of 15 is used. (R: 575078.4)

d) Determine the **price in bits to code a gray scale version** of a full image with only 128 levels of gray if the same compression factors as in c) are used. (R: 116121.6)

e) Determine the **total number of bits** that have to be spent to code only the luminance component of one image if an average number of 5 DCT coefficients are coded per block and each coefficient costs, on average, 3 bits; additionally consider that the EOB (End of Block) word costs 4 bits. How **many bits would be needed more** to also code the chrominances if an average number of 3 DCT coefficients are coded per block and each coefficient costs, on average, 2 bits (and the same EOB is used). (R: 123120, 129600 bit)

II (0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5+ 0.5 + 0.5 = 4.0 val.)

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 10 Hz at a (constant) channel bitrate of 128 kbit/s. The bits for each frame are uniformly generated in the time interval that the encoder usually dedicates to encode each image. At the encoder, the bits wait for transmission in an output buffer. Answer the following INDEPENDENT questions …

a) Assuming that the first and second frames produce 20000 bits each, what is the **minimum size of the buffer** ? (R: 14400 bit)

b) Assuming that the buffer with size 19000 has 10000 bits at 100 ms, what is the **maximum number of bits** that the second frame may produce ? (R: 21800 bit)

c) Assuming that the first frame produces 22000 bits and the buffer size is 10000 bits, what is the **maximum number of bits** that the second frame may produce ? (R: 13600 bit)

d) Assuming that the first frame produces 28000 bits and the buffer size is 20000 bits, what is the **maximum number of bits** that the third frame may produce ? (R: 30400 bit)

e) Assuming that the buffer size is 25000, what is the advisable **initial visualization delay** ? (R: 295.31 ms)

f) Assuming that the buffer size is 12000, what is **latest time** the full set of bits for the second frame may be received at the decoder ? (R: 293.75 ms)

g) What is the **minimum and maximum numbers of motion vectors per macroblock** ? (R: 0, 1)

h) What **codec module** is the main responsible for controlling the video quality ? Why ? (R: Quantizer)

i)

III (0.7 + 0.7 + 1.3 +1.3 = 4.0 val.)

Consider that your company is contacted to design a videoconference system between the various main locations of a bank. The spatial resolution is CIF (352×288 luminance samples), 4:2:0, at 12.5 Hz, with the usual number of bits per sample. Assume that you have available, offering the target video quality, two solutions:

1. **H.261 based solution** with average compression factors of 25 and 35 for the luminance and chrominance, respectively; the critical compression factors (for the images spending more bits) are 20 and 25 for the luminance and chrominances, respectively.
2. **MPEG-2 Video based solution** with N = M = 3 with average compression factors of 25 and 35 for the luminance and chrominance, respectively, for the I frames, and 30 and 45 for the luminance and chrominances, respectively, for the P and B frames. The critical compression factors are 75% of the average compression factors.

Assume that the transmission rate is always the same as the coding rate.

a) Determine the **bitrate** for the H.261-based solution. (R: 550326.9 bit/s)

b) Determine the **acquisition-visualization delay** for the H.261-based solution. (R: 103 ms)

c) Determine the **bitrate** for the MPEG-2 Video-based solution. (R: 483815.6 bit/s)

d) Determine the **acquisition-visualization delay** for the MPEG-2 Video-based solution. (R: 400 ms)

IV (1.0 + 0.5 + 1.0 + 1.0 + 0.5 = 4.0 val.)

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? (R: 117.333 kbit/s)

b) How would be a good estimation of the new **coding rate** above change if the audio bandwidth is halved ? (R: 58.667 kbit/s)

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 ? (R: Reduces due to the reducing bandwidth)

V (1.0 + 0.5 + 1.0 + 0.5 + 1.0 = 4.0 val.)

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

a) What may be **service/user implications** of the fact that a good H.264/AVC encoder 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 adopting a translational motion model ? (R: Variable block size motion compensation up to 4×4 blocks)

d) What is the **main goal** of using well selected half- and quarter-sample interpolation filters ? (R: Create better temporal predictions to reduce the prediction error and, thus, the rate for a target quality)

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