MULTIMEDIA COMMUNICATION

INSTITUTO SUPERIOR TÉCNICO, Alameda

Academic Year 2021/2022 – 1st Semester, Responsible: Prof. Fernando Pereira

1st Exam – 12th February 2022 (Saturday), 6pm

The marks should be out before **14th February (Monday), 6pm** at the CMul Web page. The exam scoring checking session will be on **16th February (Wednesday) at 2pm**, room LT4.

The exam is **2 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. ***Boa sorte !***

I (1.0 + 1.0 + 1.0 + 1.0 + 1.0 val. = 5.0 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 64 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 frame produces 15000 bits, what is the **minimum size of the buffer** ? (R: 8600 bit)

b) Assuming that the first frame produces 15000 bits and the buffer size is 18000 bits, what is the **maximum number of bits** that the second frame may produce ? (R: 15800 bit)

c) Assuming that the first frame produces 20000 bits and the buffer size is 15000 bits, what is the **maximum number of bits** that the third frame may produce ? (R: 14200 bit)

d) Assuming that the buffer size is 12800, what is **latest tim**e the full set of bits for the second frame may be received at the decoder ? (R: 400 ms)

e) Why are the **motion vectors differentially coded** ? State one positive and one negative impact of this decision.

II (1.0 + 2.0 + 1.0 + 1.0 = 5.0 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 and acquisition-visualization delay** for the H.261 based solution. (R: 550.33 kbit/s and 103 ms)

b) Determine the **bitrate and acquisition-visualization delay** for the MPEG-2 Video based solution. (R: 483.82 kbit/s and 400 ms)

c) Assuming that your client always pretends to minimize the transmission rate, **what solution** from above would you select *as a function of the minimum acquisition-visualization delay value defined by the client*?

d) State and explain two key consequences of changing N=M=3 (as above) to N=9, M=3 after describing the frame prediction structure. (R: random access reduction; compression efficiency increase; complexity increase, error resilience reduction)

III (1.0 + 1.0 + 1.0 + 1.0 + 1.0 = 5.0 val.)

Consider audio signals.

a) 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. (R: larger dynamic range)

b) Explain in what consists the **temporal post-masking effect.** What is its typical duration ? (R: 20 ms)

c) 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 ?

d) Why is the MP3 coding architecture called a **hybrid coding architecture** ?

e) Considering the varied composition of a symphonic orchestra, how and why 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 3 times the full bandwidth and ends being 2 times the full bandwidth ? (R: does not change)

IV (1.0 + 1.0 + 1.0 + 1.0 + 1.0 = 5.0 val.)

Consider the H.264/AVC video coding standard.

a) What is the offered **compression factor** if you use a H.264/AVC codec which spends around 4 Mbit/s to code video with a spatial resolution of 1152×1920, 4:2:0 at 25 Hz ? (R: 165.89)

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

c) What is the key difference between the PCM and Lossless coding modes ?

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

e) What is the key trade-off when selecting between the Intra16×16 and Intra4×4 coding modes ?