COMUNICAÇÃO DE ÁUDIO E VÍDEO

INSTITUTO SUPERIOR TÉCNICO

Year 2017/2018 – 1st Semester, Responsible: Prof. Fernando Pereira

2nd Exam – 29th January 2018, 8am (Monday)

The marks should be out before **13th February (Tuesday), 8pm** at the CAV Web page and the exam checking session will be on the **14th February (Wednesday), 2pm** in room LT4.

The exam is **3 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 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 val. = 3.5 val.)

Consider a digital transmission system.

a) What is a main difference in terms of rate production between source coding and channel coding ? Why ?
b) Under what conditions may a channel coding related 'bit' typically add quality to an image ? (R: when there are errors in the channel)
c) What may be the impact of the Weber's Law when coding an image ? Why ?
d) What is the first major impact of the limited human capacity to see spatial detail when coding an image ? Why?
e) What is the first major impact of the lower human sensibility to color in comparison with luminance when coding an image ? Why?
f) What may be a major impact of the limited human capacity to see temporal changes when coding video ? Why?

II (1.0 + 1.0 + 1.0 + 1.0 val. = 4.0 val.)

Consider the JPEG standard to code photographic images with a 288×352 luminance resolution and 8 bit/sample.

a) Determine the 'price' in bits per pixel if 4:4:4, 4:2:2 and 4:2:0 color subsampling is used and no compression is used. (R: 24, 16, 12 bpp)
b) Determine the average 'price' in bits per pixel if 4:4:4, 4:2:2 and 4:2:0 color subsampling is used and compression is applied with a luminance compression factor of 10 and a chrominances compression factor of 15. (R: 1.866, 1.333, 1.066 bpp)
c) Determine the average 'price' in bits for one single luminance sample AND for one single chrominance sample if compression is applied with a luminance compression factor of 15 and a chrominances compression factor of 20. (R: 0.5333 and 0.4 bps)
d) Determine the total number of bits that have to be spent to code all the components of a full 4:2:2 image if, on average, 3 DCT coefficients are coded per block and each DCT coefficient costs, on average, 4 bits; additionally consider that the EOB (End of Block) word costs 4 bits. (R: 50688 bit)

III (0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 = 3.0 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 12.5 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 frame produces 30000 bits, what is the minimum size of the buffer ? (R: 19760 bit)

b) Assuming that the first frame produces 25000 bits and the buffer size is 18000 bits, what is the maximum number of bits that the third frame may produce ? (R: 23720 bit)

c) Assuming that the buffer size is 18000 bits, what is the maximum number of bits that the first frame may produce ? (R: 28240 bit)

d) Assuming that the first frame produces 20000 bits and the buffer size is 15000 bits, what is the maximum number of bits that the second frame may produce ? (R: 15480 bit)

e) Assuming that the buffer size is 18000, what is latest time the full set of bits for the first frame may be received at the decoder ? (R: 220.625 ms)

f) Assuming the second frame may be displayed at 320 ms without delay related problems, what is the minimum size of the buffer ? (R: 20480 bit)

IV (1.0 + 1.5 + 1.0 = 3.5 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: 103 ms)

b) Determine the bitrate and acquisition-visualization delay for the MPEG-2 Video based solution. (R: 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 acquisition-visualization delay requirement defined by the client ?

V (0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 = 3.0 val.)

Consider audio signals.

a) Indicate two audio characteristics which are numerically larger than the corresponding speech characteristics. (R: bandwidth and dynamic range)

b) What is the main audio characteristic justifying to use a higher bit depth than for speech ? (R: higher dynamic range)

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

d) What is the main reason to increase the size of the audio frames in MPEG-1 Audio Layer 2 regarding Layer 1?

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 increasing the sampling rate from a value which starts being 2.5 times the full bandwidth and ends being 4 times the full bandwidth?

VI (0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 = 3.0 val.)

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 the so-called motion to photon delay ?

f) What is the most relevant parameter the user can (indirectly) control when navigating the visual data and which is critical to select the visual data to be displayed ? (R: viewing direction)