COMUNICAÇÃO DE ÁUDIO E VÍDEO

INSTITUTO SUPERIOR TÉCNICO

Year 2018/2019 – 1st Semester, Responsible: Prof. Fernando Pereira

1st Exam – 11th January 2019, 8am (Friday)

The marks should be out before **14th January (Monday), 6pm** at the CAV Web page and the exam checking session will be on the **15th January (Tuesday), 10am** 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 (0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 val. = 3.0 val.)

a) Why is the user important in a multimedia service (2 reasons) ?

b) Why is it fundamentally different the impact of noise in an analogue and in a digital signal ?

c) If you had to find a simple way to estimate how much more complex is an audio signal than a speech signal what would you propose ?

d) Why may it be reasonable to state that a microphone and a loudspeaker are 'two sides of the same coin' ?

e) Which one is more important for audio coding, the threshold of hearing or the threshold of feeling ? Why ?

f) Why is visual acuity important for video services ?

II (1.0 + 0.5 + 0.5 + 0.5 + 1.0 val. = 3.5 val.)

Consider the JPEG standard to code photographic images with a 1152×1440 luminance resolution, 4:2:0 color subsampling and 8 bit/sample.

a) How many pixels, samples and blocks exist in this type of image. (R: 1658880, 2488320, 38880)

b) Determine the price in bits per pixel if a codec with a luminance compression factor of 25 and a chrominances compression factor of 15 is used. (R: 0.5867)

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: 973209.6)

d) Determine the price in bits to code a gray scale version of a full image with only 64 levels of gray if the same codec as in c) is used. (R: 398131.2)

e) Determine the total number of bits that have to be spent to code the 2 chrominance components of an 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. (R: 246240)

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 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 and second frames produce 10000 bits each, what is the minimum size of the buffer ? (R: 7200)

b) Assuming that the buffer size is 18000 bits, what is the maximum number of bits that the first frame may produce ? (R: 24400)

c) Assuming that the first frame produces 12000 bits and the buffer size is 10000 bits, what is the maximum number of bits that the second frame may produce ? (R: 10800)

d) Assuming that the first frame produces 12000 bits and the buffer size is 10000 bits, what is the maximum number of bits that the third frame may produce ? (R: 16400)
e) Assuming that the buffer size is 18000, what is the advisable initial visualization delay ? (R: 381.25 ms)
f) Assuming that the buffer size is 12000, what is latest time the full set of bits for the first frame may be received at the decoder ? (R: 287.5 ms)

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

Assume that you are contacted by a company to design a digital storage system for video clips. The company requires some editing flexibility and needs to store the largest number of 4 minutes clips in a disk. The maximum access speed to the disk is 80 Mbit/s. The clips have 4K resolution with the following characteristics: 3840 x 2160 (Y), 4:2:2, 10 bit/sample at 25 Hz.

a) Assuming that you have at your disposal, providing the required video quality, a JPEG coding solution with average compression factors of 40 and 45 for the luminance and chrominances, respectively, determine the maximum access time for an image knowing that the compression factors for critical frames are 20% lower than average. (R: 61.12 ms)
b) Assuming now that you have at your disposal, providing the required video quality, a MPEG-2 Video coding solution with N=12 and M=4 with the following average 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

   Determine the maximum access time for an image knowing that the compression factors for critical frames are 25% lower than average. (R: 346.26 ms)
c) Determine, justifying, which coding solution would you propose to your client if the target is only to maximize the number of clips stored in the disk. (R: MPEG-2)
d) Determine, justifying, which coding solution would you propose to your client if a maximum random access requirement of 100 ms is put forward together with the requirement of maximizing the number of clips stored in the disk. (R: JPEG)
e) How many full video clips would you be able to store in the disk for the JPEG solution if the disk has a capacity of 10 TByte (10^{12}). (R: 3404)

V (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 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 chrominance, 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 chrominance, 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: 550326.9 bit/s, 103 ms)
b) Determine the bitrate and acquisition-visualization delay for the MPEG-2 Video based solution. (R: 483815.6 bit/s, 400 ms)
c) Assuming that your client always pretends to minimize the transmission rate, what solution from above would you select depending on the acquisition-visualization delay requirement ?

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

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 ?