SOLUTIONS

Module:	Multimedia Fundamentals		
Module Code	EBU5303	Paper	A
Time allowed	2hrs	Filename	**Solutions_2021_EBU5303_A**
Rubric	ANSWER ALL FOUR QUESTIONS		
Examiners	Dr Marie-Luce Bourguet	Dr Atm Shafiul Alam	

Question 1

a) This question is about digitisation.

[10 marks]

i) You want to digitise the complex waveform of a sound. What is aliasing and what do you need to know about the waveform to avoid aliasing?

(4 marks)

Solution: Aliasing is equivalent to sampling errors (or under sampling), the effect of aliasing is distortion of the sound and loss of quality. To avoid aliasing, we need to know the sound's frequency components by Fourier analysis.

ii) Still considering the waveform of question i) above, what sampling rate do you then use to avoid aliasing?

(2 marks)

Solution: The sampling rate should be twice the highest frequency in the sound, according to the Nyquist theorem.

iii) Calculate the size of a video file, in bytes, which has the following characteristics: the pixel dimension is 200 pixels x 300 pixels, true colour, the frame rate is 24 frames/s, the audio track contains speech digitised at 15 kHz and 8 bits/sample, the duration is 1 minute. Show the details of your calculations.

(4 marks)

Solution: Visual track size = 200x300x24x24x60/8 bytes = 259.2 Mbytes.
Audio track size = 15000x8x60/8 bytes = 0.9 Mbytes
Total size = visual + audio = 259.2 + 0.9 = 260.1 Mbytes

b) This question is about colour encoding.

[15 marks]

i) Explain why a grayscale image is three times smaller (i.e. contains three times less data) than a true colour image.

(3 marks)

Solution: A true colour image encodes each pixel using 24 bits, a grayscale image encodes each pixel using 8 bis. The bit depth of grayscale is thus 3 times smaller than the bit depth of true colour.

ii) Describe the properties of an unsaturated colour.

(3 marks)

Solution: An unsaturated colour contains some white/grey, which means that it contains many different wavelength components, or that it isn't pure.

iii) How is grey represented in RGB? How about CMY? And how about HSV?

(3 marks)

Solution: In RGB, grey is represented by equal amounts of R, G and B; in CMY, grey is also represented by equal amounts of C, M and Y; in HSV, grey is represented by S=0.

iv) In the (C, M, Y) model, explain how you can decrease the saturation of a colour.

(3 marks)

Solution: In CMY, saturation is decreased by reducing the values of C, M and Y, i.e., by putting less ink on paper which will reduce their absorption power and let some white light be reflected and mixed with the colour.

v) What (H, S, V) values would you use to encode an unsaturated dark red colour?

(3 marks)

Solution: H = 0; S is between 1 and 99 (or between 0.1 and 0.9); V is between 1 and 50 (or 0.1 and 0.5).

Question 2

a) Devise a process to transform a true colour image into a binary image (open ended question).
[5 marks]

Solution: Possible solution: Binary means that a pixel is assigned 0 or 1. We could convert RGB to YUV; all "dark" pixels, i.e., Y<127 are assigned 0; all "bright" pixels are assigned 1.

Other solutions possible …

b) This question is about audio.
[10 marks]

i) What type of speech sound is used for doing speech recognition and why?
(5 marks)

Solution: The voiced sounds are semi-periodic and characterised by their formants which are characteristic frequencies indicating particular sounds such as vowels, useful for speech recognition.

ii) Speech is usually encoded using a smaller bit depth than music. Explain why and give typical bit depths for speech and for music.
(5 marks)

Solution: Speech doesn't have to be encoding with great details to remain comprehensible, as speech poor signal quality can be compensated by prior knowledge about the spoken language. At contrary, as many details as possible should be encoded for music to remain enjoyable . Typical bit depth for speech is 8, for music it is 16.

c) This question is about image lossless compression.
[10 marks]

i) What image properties are used in RLE encoding to achieve compression? Justify your answer.
(3 marks)

Solution: The image must contain lots of spatial redundancy, i.e. long sequences of pixels of the same colour.

ii) You are compressing a true colour image of the Chinese flag using RLE. The pixel dimension is 1600x1200 pixels and the first row is entirely red. How many bytes are needed to encode the first row if you consider that it contains exactly one run? Show your calculations.
(4 marks)

Solution: The run is 1600 pixel long. So the maximum "size" to be encoded is 1600, which requires 11 bits to be encoded, or 2 bytes. The colour red requires 3 bytes in RGB, so in total, we need 5 bytes.

iii) Recalculate the number of bytes needed to encode the first row of the image of question ii) above, but this time the bit depth for encoding the size is fixed to one byte.

(3 marks)

Solution: The row must be split into several runs of maximum length 256. 1600 pixels are organised into 7 runs of 4 bytes each, 1 for the size and 3 for the colour. The total will be 7x4 = 28 bytes.

Question 3

a) This question is about JPEG.

[15 marks]

i) JPEG is lossy, but some JPEG images look exactly like the original image after decompression. How is this possible?

(3 marks)

Solution: It is possible because JPEG has only removed the details that are not perceptible, i.e., the perceptual redundancy, the JPEG image is smaller than the original one, but they are perceptually undistinguishable.

ii) How is the quantisation factor used to influence the compression rate?

(3 marks)

Solution: The quantisation factor is used to multiple the quantisation table, i.e., it increases or decreases the quantisation table's values, influencing the compression rate as, for example, higher values in the table will increase the compression rate.

iii) During the pre-processing step, images are converted from RGB to YC_bC_r. Why?

(3 marks)

Solution: To separate luminance and colour information for each pixel, which makes chroma-subsampling possible, i.e. the removal of some of the colour information.

iv) How does DPCM work and why is it applied to the DC coefficients?

(3 marks)

Solution: DPCM encodes the difference between two values rather than the values themselves. It is applied to the DC coefficients because the DC coefficients of two successive blocks are likely to be similar, hence their difference will be small and necessitate few bits for encoding.

v) To help image decompression, what information do you think should be stored in the JPEG image file header?

(3 marks)

Solution: The quantisation tables, the quantisation factor, the chroma-subsampling scheme, the Huffman tables.

b) This question is about MPEG.

[10 marks]

i) MPEG exploits three types of redundancy: spatial, perceptual and temporal. Briefly explain how each of them is used in MPEG.

(6 marks)

Solution: Spatial redundancy when RLE encoding is applied to the AC coefficients after zigzag scanning.

Perceptual redundancy during quantisation when the AC coefficients are rounded to zero. Temporal redundancy during motion estimation when looking for matching macroblocks between consecutive frames.

ii) Is the Block Matching Algorithm (BMA) an example of fast motion estimation? Justify your answer.

(4 marks)

Solution: No, BMA simply divides a frame into macroblocks and scan the previous frame within a search window to find a matching macroblock, it is slow and computationally expensive. Fast motion estimation tries to optimise this process by reducing the number of search points.

Question 4

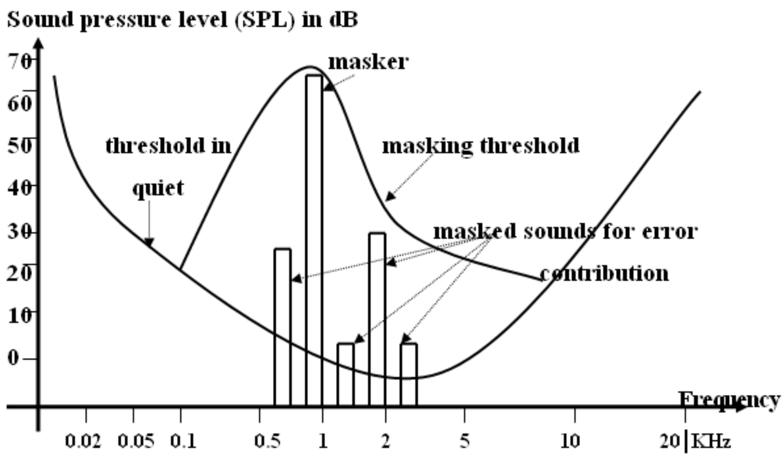
a) This question is about perceptual encoding and MP3.

[15 marks]

i) Explain how frequency masking modifies the threshold of hearing. Illustrate your answer with a drawing.

(5 marks)

Solution: Frequency masking raises the threshold of hearing, because more volume becomes necessary to hear the masked tones, which are of similar or higher frequency than the loud tone (the masker tone).



for the shape of the threshold in quiet, for the shape of the masking threshold.

ii) Let us assume that an uncompressed band value is 20,000 and values from all bands are quantised by dividing by 128 and rounding down. What is the quantisation error? Show your calculations.

(3 marks)

Solution: 20,000/128 = 156 is the quantised value. 156*128 = 19,968. (20,000-19,968)/20,000 = 0.0016 is the quantisation erro.

iii) Now suppose that this band requires less precision because of a strong masking tone, and that it should be scaled by a factor of 0.3. Recalculate the quantisation error.

(3 marks)

Solution: 20,000*0.3/128 = 46 is the new quantised value. 46*128/0.3 = 19,627. (20,000-19,627)/20,000 = 0.019 is the new quantisation error.

iv) With an MP3 bitrate of 128 kbit/s, calculate the compression ratio that is achieved on a speech audio signal sampled at 20KHz and with a bit depth of 8.

(2 marks)

Solution: uncompressed birate: 20,000x8 = 160 kbit/s. Compression ratio: 160:128 = 5:4.

v) Is the human ear more sensitive to low or high frequencies? Justify your answer by referring to what you know about the critical bands in the human hearing range.

(2 marks)

Solution: The human ear is more sensitive to low frequencies. The critical bands in the low frequencies are narrower than in the high frequencies.

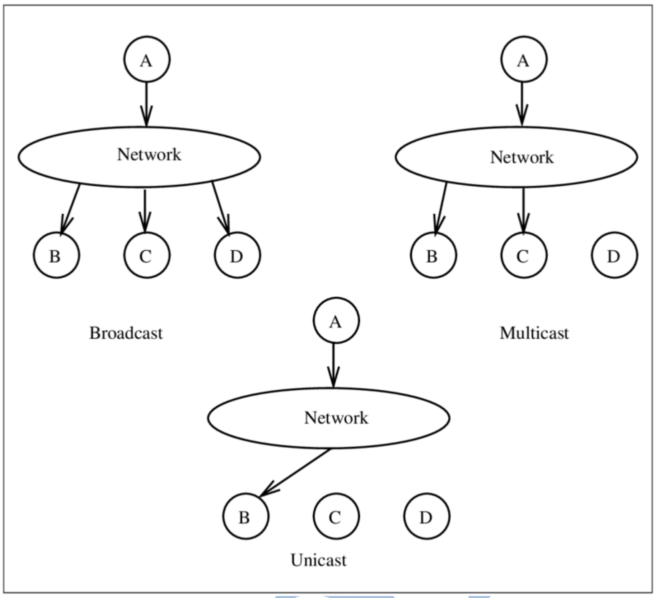
b) This question is about broadcasting.

[10 marks]

i) Explain the difference between unicast, broadcast and multicast. Illustrate your answer with a drawing.

(4 marks)

Solution: With unicast, there is a single sender and a single recipient. With broadcast, a sender transfers packets to all potential receivers. With multicast, a sender transfers packets to receivers who requested it.



ii) Briefly explain the following terms: source coding, transport stream and channel coding. Give examples.

(6 marks)

Solution: source coding refers to the compression of the digital signal, e.g., MPEG2; the transport stream specifies a container for the multiplexed signals, e.g., MPEG2TS; channel coding is a process of detecting and correcting bit errors, e.g., Reed-Solomon.