SOLUTIONS

Module:	Multimedia Fundamentals		
Module Code	EBU5303	Paper	A
Time allowed	2hrs	Filename	Solutions_201920_EBU5303_A
Rubric	ANSWER ALL FOUR QUESTIONS		
Examiners	Dr Marie-Luce Bourguet	Dr Atm Shafiul Alam	

Question 1

a) This question is about digitisation. Consider a sound wave W with a frequency f = 440 Hz.
[8 marks]

i) What is the sine function representing W? **(1mark)**

s (t) = sin (880πt)

ii) What kind of sound is represented by a completely regular sine wave such as W? **(1 mark)**
A pure tone

ii) What does the amplitude of W tell us about the sound it represents? **(1 mark)**

Its loudness

iii) What is the minimum sampling rate you should use to ensure that you can digitise W without audio aliasing? Justify your answer. **(2 marks)**

880 Hz, following the Nyquist theorem

iv) You decide to use 5 bits per sample. How many different values can W take? **(1 mark)**

$2^5 = 32$

v) Calculate an approximation of the Signal-to-Quantisation Noise Ratio (SQNR) of W. Explain your calculation. **(2 marks)**

SQNR ~= 6n where n is the number of bits/sample = 30 dB

b) This question is about colour encoding.
[12 marks]

i) In a true colour image, what is the number of different colours that can be represented? Justify your answer. **(2 marks)**

True colour means 24 bits colour depth. Number of colours is 2^{24} ~= 16.7 million.

ii) Describe the properties of a fully saturated colour. **(2 marks)**

It is not mixed with white, i.e. it contains a narrow range of wavelengths

iii) In the HSV colour model, how is the grayscale represented? **(2 marks)**

Along the vertical V axis. Black is at the bottom, white at the top.

iv) What (R, G, B) values would you use to encode an unsaturated dark blue colour? **(2 marks)**

for example (20, 20, 100)

v) What (C, M, Y) values would you use to encode a fully saturated bright green colour? **(2 marks)**

(255, 0, 255)

vi) Yellow ink is spread onto a white sheet of paper. What colour will you see if the paper is illuminated with a blue light? Justify your answer. **(2 marks)**

Yellow absorbs blue, so the paper will appear black.

c) Consider a video with the following properties: frame size is 1280-by-720-pixels; colour depth is 24-bits; frame rate is 30 fps; duration is 1 minute.

[5 marks]

i) How much data rate reduction can be achieved by reducing the width and height of this video to half? Prove your answer by calculating the data rates. **(3 marks)**

The file size will be reduced to a quarter of the original
Original size : 1280*720*24*30 = 663,552,000 bits/second
Reduced size: 640*360*24*30 = 165,888,000 bits/second = 663,552,000 / 4

ii) Give two possible strategies for reducing the colour depth of the video to just a third of its original depth.

Convert the images to grayscale or use a 256 entries (8 bits) CLUT.

Question 2

a) This question is about audio.

[5 marks]

i) Briefly explain what is shown in an audio histogram such as the one in Figure 1. In particular, comment the units used on the X and Y axes.

It shows how many samples there are at each amplitude level in the audio selection. X axis shows amplitude in decibels; Y axis shows number of samples.

ii) Do audio histograms and audio spectrograms represent signals in the same domain? Justify your answer.

No, audio histograms are in the temporal domain; audio spectrograms are in the frequency domain.

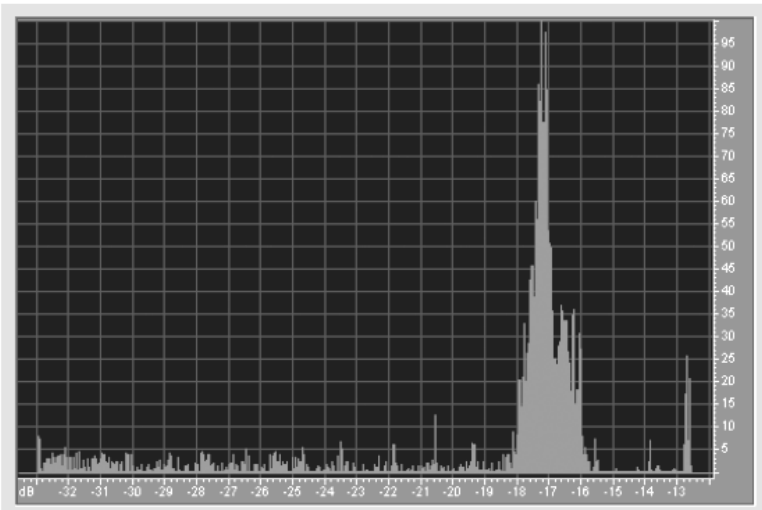


Figure 1**: Audio histogram**

b) This question is about image lossless compression.

[10 marks]

i) What image property is used in Huffman encoding to achieve compression? Justify your answer.

Statistical (entropy) property, i.e. the fact that some colours appear more often than others. The frequent colours are coded using short codes.

ii) Decode the following binary message using the Huffman encoding table provided in Table 1.

Binary message:
1111101111010011011110101101100110111110101011001100100111111

EDECCBCECCDDBCEDCCDBCBAEE

iii) Assuming that each symbol of Table 1 would normally be encoded using 3 bits (which is enough to encode 5 different symbols), how much compression is achieved in the binary message of question ii) above?

Size of the binary message: 61 bits; size of the decoded message 25x3 = 75 bits; compression is 75/61 = 1.23

iv) Consider the following statement: “For compression to remain lossless, an image should be encoded/decoded only once”. Is it correct? Justify your answer.

No, with lossless encoding, images can be encoded/decoded many times, as the decoded image is the same as the original one.

symbol	probability	code
A	0.40	00
B	0.20	01
C	0.20	10
D	0.10	110
E	0.10	111

Table 1: Huffman encoding table

c) Consider the block diagram shown in Figure 2 and answer the questions below.

[10 marks]

i) Describe what is contained inside the block marked with the label “A”.

8x8 blocks of pixels which have been converted from RGB to YUV and chroma sub-sampled

ii) What step of the JPEG compression process is marked with the label “B”? Explain this step.

Discrete Cosine Transform (DCT). DCT transforms the image from the spatial to the frequency domain to obtain 8x8 blocks of DCT coefficients. The coefficients are arranged from lowest frequency (average value for the block) to highest frequency, i.e. fine details.

iii) Explain the use of quantisation tables in the JPEG compression process.

The 8x8 blocks of DCT coefficients are divided by these quantisation tables. Different quantization tables are used according to the amount of compression that is needed. High quantisation values achieve greater compression but may incur a loss of quality.

iv) Which phenomenon of human vision is exploited in chroma sub-sampling?

The human eye is more sensitive to changes of luminance (brightness) than it is to changes of chrominance (colour)

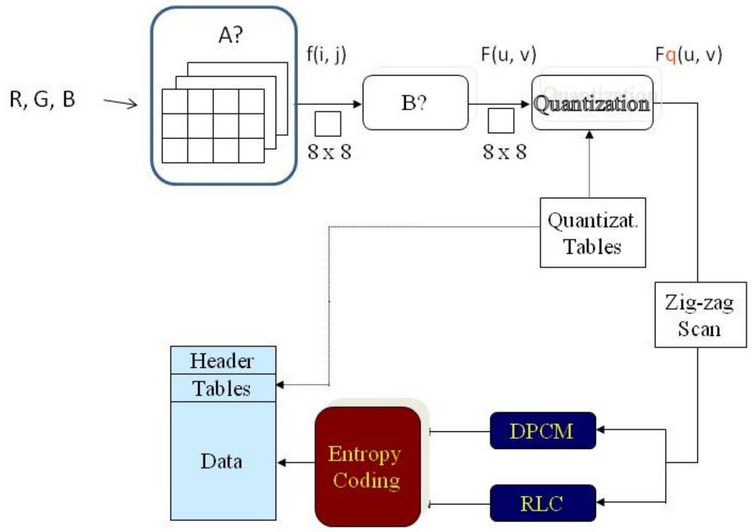


Figure 2: JPEG block diagram

Question 3

a) This question is about MPEG.

[14 marks]

i) What type of MPEG frame makes no prediction? **(1 mark)**

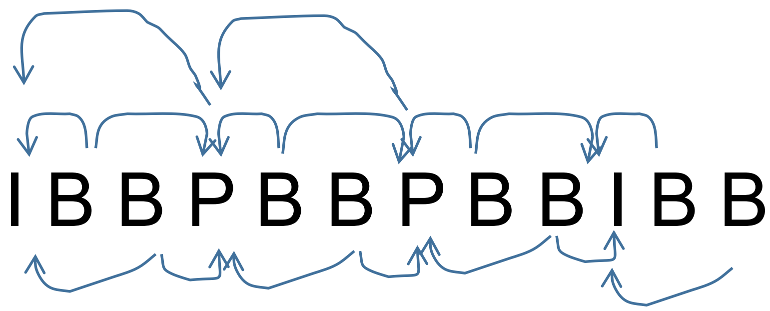
Solution: I-frames

ii) What type of MPEG frame is never used as a reference frame during decompression? Why? **(3 marks)**

Solution: B-frames, because they are highly predicted and prediction errors would be propagated to other frames.

iii) Suppose an MPEG encoder uses the nine-frame sequence IBBPBBPBB. Draw a diagram showing the dependencies between the <u>first 12 frames</u> of a compressed clip produced by this encoder. **(3 marks)**

Solution: for choice of 12 frames; for correct dependencies shown; for all dependencies shown.



iv) Briefly explain the Block Matching Algorithm (BMA) for motion estimation when a maximum motion displacement of x pixels is used. **(4 marks)**

Solution: A frame is divided into macroblocks of 16x16 pixels. Macroblocks are matched against corresponding blocks in the previous or following frame within the square window of

width (16 + 2w) pixels. A function such as the Mean Squared Error (MSE) function (matching criterion) is used to find the best match.

v) Given a maximum motion displacement of 8 pixels, how many evaluations of the matching criterion are required in the BMA? **(1 mark)**

Solution: $(2w+1)^2 = 17^2$

vi) Briefly explain the basic principle of fast motion estimation techniques. **(2 marks)**

Solution: the number of search points is reduced by selectively checking only a small number of specific points.

b) This question is about perceptual encoding.

[11 marks]

i) With A-law coding, larger signals are represented with greater precision – more data bits – than smaller signals. Is this statement true or false? Justify your answer.

(4 marks)

Solution: false, it is the opposite. For the human auditory system, high amplitude sounds do not require the same resolution as low amplitude sounds, i.e. the human ear is more sensitive to quantisation noise in small signals than large signals, so A-law encoding uses more bits for small amplitude signals.

ii) There are 24 critical bands in the human hearing range, but critical bands for low frequencies are narrower than those for high frequencies. What is this statement telling us about the human ability to distinguish between frequencies?

(3 marks)

Solution: Human ability to distinguish between frequencies decreases nonlinearly from low to high frequencies. At the very lowest audible frequencies, we can tell the difference between pitches that are only a few Hz apart, while at high frequencies the pitches must be separated by more than 100 Hz before we notice a difference.

iii) What is the threshold of hearing and how does frequency masking affect the threshold of hearing?

(4 marks)

Solution: the threshold of hearing is the minimal level at which a sound can be heard, it depends on the frequency of the sound. Frequency masking causes the threshold of hearing to be raised within a critical band in the presence of a masking tone.

Question 4

a) This question is about MP3.

[15 marks]

i) In MP3, one way to reduce the amount of data in the compressed signal is to use scaling factors that increase the quantisation error where it doesn't matter. Briefly explain how the parts of the signal that will be multiplied by a large scaling factor can be found. **(5 marks)**

Solution: The psychoacoustical analysis module outputs a set of signal-to-mask ratios (SMRs). Bands that have a low SMR are multiplied by larger scaling factors because the quantisation error for these bands has less impact, falling below the masking threshold.

ii) Say that an uncompressed band value is 10,000 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: 10,000/128 = 78 is the quantised value. 78*128 = 9,984.
(10,000-9,984)/10,000 = 0.0016 is the quantisation error.

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 0.1. Recalculate the quantisation error. **(3 marks)**

Solution: 10,000*0.1/128 = 7 is the new quantised value. 7*128/0.1 = 8,960
(10,000-8,960)/10,000 = 0.104 is the new quantisation error.

iv) With an MP3 bitrate of 128 kbit/s, calculate the compression ratio that is achieved on a CD quality digital audio signal (CD quality = 44100 samples per second, stereo and 16 bits per channel). **(2 marks)**

Solution: uncompressed birate: 44100x32 = 1411 kbit/s. Compression ratio: 1411:128 ~= 11:1

iv) What is meant by "Average Bit Rate" (ABR)? **(2 marks)**

Solution: Average Bit Rate (ABR) is a type of Variable Bit Rate (VBR) where the bitrate is allowed to vary for more consistent quality, but is controlled to remain near an average value chosen by the user, for predictable file sizes.

b) This question is about DVB-S.

[10 marks]

i) What compression standard is used for source coding in DVB-S? **(2 marks)**

Solution: MPEG2.

ii) What is the purpose of energy dispersal? **(2 marks)**

Solution: it is used to achieve a power-density spectrum of the modulated signal that is as even as possible.

iii) How does the Reed-Solomon Error Protection scheme work? **(2 marks)**

Solution: Parity bits are introduced in each transport packet.

iv) Assuming a symbol rate of 27.5 MS/s, QPSK modulation, Reed-Solomon code with rate (204, 188), and a code rate of ¾ are used, calculate the bit stream net data rate. Show your calculations. **(4 marks)**

Solution: gross_data_rate = 2 bits/symbol (QPSK) x 27.5 Megasymbols/s = 55 Mbit/s

Net data rate (Reed-Solomon) = gross_data_rate x 188/204 = 50.69 Mbit/s

Net data rate = code rate x net data rate (Reed-Solomon) = 3/4 x 50.69 = 38.02 Mbit/s