CARDIFF UNIVERSITY
EXAMINATION PAPER

Academic Year:	2010/2011
Examination Period:	Autumn
Examination Paper Number:	CM0340 Solutions
Examination Paper Title:	Multimedia
Duration:	2 hours

Do not turn this page over until instructed to do so by the Senior Invigilator.

Structure of Examination Paper:

There are 11 pages.
There are 4 questions in total.
There are no appendices.
The maximum mark for the examination paper is 81 and the mark obtainable for a question or part of a question is shown in brackets alongside the question.

Students to be provided with:

The following items of stationery are to be provided:
ONE answer book.

Instructions to Students:

Answer 3 questions.
The use of calculators is permitted in this examination.
The use of translation dictionaries between English or Welsh and a foreign language bearing an appropriate departmental stamp is permitted in this examination.

Q1. (a) *What are the differences between analog signals and digital signals?*

Analog Signals — continuous signals of some time varying quantities; can't be processed directly by computers [1]

Digital Signals — digital samples of the signals at regular interval; can be readily processed by computers [1]

2 Marks — Bookwork

A computer is to be used to add effects to analog audio signals. What two *types of devices in general are needed? Describe their functionalities in the processing pipeline.*

Analog-to-Digital Converter (ADC) : take analog signals from analog sensor and digitally sample data. [1]

Digital-to-Analog Converter (DAC) : take digital signals from computer and outputs an analog signal that may be displayed by output device. [1]

2 Marks — Bookwork

(b) *Audio signals are often sampled at different rates. CD quality audio is sampled at 44.1kHz rate while telephone quality audio sampled at 8kHz. What are the maximum frequencies in the input signal that can be fully recovered for these two sampling rates? Briefly describe the theory you use to obtain the results?*

- CD quality audio, the maximum frequency: 44,100Hz / 2 = 22,050Hz. [1]
- Telephone quality audio, the maximum frequency: 8kHz / 2 = 4kHz. [1]
- This is based on *Nyquist theorem*: the sampling frequency for a signal must be at least twice the highest frequency component in the signal. [1]

3 Marks — Unseen problem based on theories covered in lectures

If an arbitrary input signal is directly sampled, what artefact may result and how to solve this?

This may result in *aliasing* artefact. [1]

To solve this, add an *analog low pass filter* before sampling to eliminate high frequency components. [1]

2 Marks — Bookwork

(c) *Using wavetable for digital audio synthesis, for two audio samples as follows (simplified for ease of calculation)*

$S_1 : 0, 1, 0, -1, 0, 1, 0,$ $S_2 : -1, -0.6, -0.4, 0, 0.4, 0.6, 1,$

each with 7 samples (1–7). What is the output of linear crossfading from S_1 to S_2, assuming samples 3–5 are transitional samples with mixed information from both sources?

The envelope for S_1 is $E_1 = (1, 1, 0.75, 0.5, 0.25, 0, 0)$ and the envelope for S_2 is $E_2 = (0, 0, 0.25, 0.5, 0.75, 1, 1)$, calculate $S = S1. * E1 + S2. * E2$ (elementwise multiplication), we have $0, 1, -0.1, -0.5, 0.1, 0.6, 1$ [4]

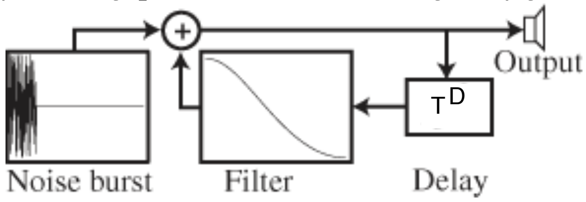
For a wavetable entry S_3 representing a short audio sample with fundamental frequency of 600Hz. From this, a reshaped wave S_4 is derived satisfying $S_4(t) = S_3(0.75t)$. What is the fundamental frequency of S_4 assuming the same sampling rate is used for playback?

Assume the period of the original audio is T. Since $S_4(\frac{4}{3}T) = S_3(T)$, the period of the new audio is $\frac{4}{3}T$. The frequency of the new audio is $\frac{3}{4}f = \frac{3}{4} \times 600 = 450$Hz.

[2]

6 Marks — Unseen problem

(d) *Answer the following questions based on the given figure:*



i. *What is the audio synthesis algorithm described in the diagram?*
This diagram illustrates *Karplus-Strong* audio synthesis algorithm. [1]

ii. *Which general approach of digital audio synthesis most accurately does this algorithm belong to?*
The algorithm belongs to *Physical Modelling* approach. [1]

iii. *If the audio is sampled at 44.1kHz, and $D = 100$ (see the figure), what is the fundamental frequency of the synthesised audio?*
According to the relationship

$$D = \frac{F_s}{F_1},$$

where D is the delay, F_1 is the fundamental frequency and F_s is the sampling frequency. We have $F_1 = \frac{F_s}{D} = \frac{44100}{100} = 441Hz.$ [2]

iv. *To halve the fundamental frequency of the synthesised audio, what* two *possible changes can you make?*
According the relationship, to halve the fundamental frequency F_1, either halve sampling frequency F_s or double the delay D. [2]

6 Marks – Unseen problem with the theories covered in lectures

(e) *Describe what ADSR envelope means (with an illustration) and how this is applied in digital audio synthesis.*

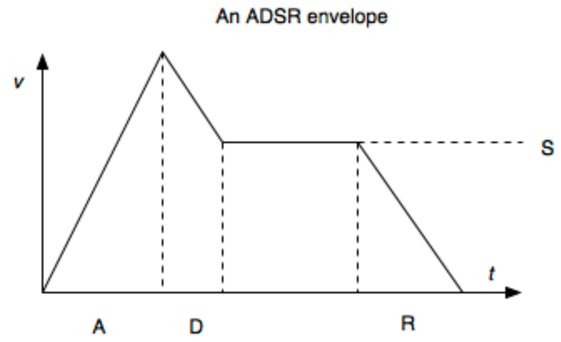
ADSR means

Attack : How quickly the sound reaches full volume after the sound is activated. [1]

Decay : How quickly the sound drops to the sustain level after the initial peak. [1]

Sustain : The "constant" volume that the sound takes after decay until the note is released. [1]

Release : How quickly the sound fades when a note ends. [1]



(illustration) [1]

ADSR is used to modulate some aspects of the instrument's sound often its volume over time, simulating the behaviour of mechanical instruments. [1]

6 Marks – Bookwork

Question 1 Total Marks 27

Q2. (a) *Dithering is often used when converting greyscale images to monochrome.*

i. *What is the basic idea of dithering?*
Dithering is to use a larger pattern to approximate the (greyscale) levels of the input image. [1]
1 Mark — Bookwork

ii. *For the given* 2×2 *dither matrix, briefly describe the* ordered dithering *algorithm.*

$$\begin{pmatrix} 0 & 2 \\ 3 & 1 \end{pmatrix}$$

The algorithm involves the following steps

- re-map each intensity to a range of 0 to 4 by dividing the intensity (for 256 levels) by (256/5).
- tile the dither matrix to the same dimension as the input image.
- if the remapped intensity is larger than the corresponding dither matrix entry, put a 1 otherwise 0

[2]

2 Marks — Bookwork

iii. *Use the same dither matrix, what is the result for the following input? Assume that the input is greyscale intensities normalised to* **0 to 1**.

$$\begin{matrix} 0.66 & 0.18 & 0.03 & 0.19 \\ 0.54 & 0.13 & 0.56 & 0.37 \\ 0.70 & 0.99 & 0.88 & 0.46 \\ 0.67 & 0.17 & 0.67 & 0.98 \end{matrix}$$

The remapped intensities are calculated by $floor(I * 5)$ for I being the input intensity.

$$\begin{matrix} 3 & 0 & 0 & 0 \\ 2 & 0 & 2 & 1 \\ 3 & 4 & 4 & 2 \\ 3 & 0 & 3 & 4 \end{matrix}$$

The tiled dither matrix

$$\begin{matrix} 0 & 2 & 0 & 2 \\ 3 & 1 & 3 & 1 \\ 0 & 2 & 0 & 2 \\ 3 & 1 & 3 & 1 \end{matrix}$$

Compare each element of A with B, the final result:

$$\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}$$

[4]

4 Marks — Unseen problem

(b) *Different colour models are often used in different applications. What is the CMYK colour model? Give an application in which this colour model is mostly used and explain the reason.*

The CMYK colour model use Cyan, Magenta, Yellow and Black as primaries (components). [1]

The CMYK colour model is mostly used in printing because the colour pigments on the paper absorb certain colours thus a subtractive model is suitable; black is used to produce darker black than simply mixing CMY. [2]

3 Marks — Bookwork

Given a colour represented in RGB colour space as $R = 0.2$, $G = 0.6$, $B = 0.3$, what is its representation in the CMYK colour model?

First convert to CMY as

$$\begin{pmatrix} \bar{C} \\ \bar{M} \\ \bar{Y} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.4 \\ 0.7 \end{pmatrix}$$

Then

$$K = min(\bar{C}, \bar{M}, \bar{Y}) = 0.4,$$
$$C = \bar{C} - K = 0.4,$$
$$M = \bar{M} - K = 0,$$
$$Y = \bar{Y} - K = 0.3.$$

[2]

2 Marks — Unseen problem

(c) *Give* three *colour models other than RGB/CMYK and explain the benefits of using the model by showing a practical application for each model.*

A possible answer:

CIE L*a*b* : relate more closely to human perception, useful for image processing such as Photoshop [2]

YUV(YCbCr) : separate colour from luminance component, useful for PAL video/MPEG [2]

YIQ : separate colour from luminance component, useful for NTSC video/JPEG [2]

Other sensible answers are acceptable as well.

6 Marks — Bookwork, 1 mark for each colour model and 1 mark for each application

(d) *What is chroma subsampling? Why is chroma subsampling meaningful? What is the benefit of doing chroma subsampling?*

Chroma subsampling is a method that stores colour information at lower resolution than intensity information. [1]

Chroma subsampling is meaningful because human visual system is less sensitive to variations in colour than brightness. [1]

Chroma subsampling can reduce the bandwidth for colour detail in almost no perceivable visual difference. [1]

3 Marks — Bookwork

For the following array of colour values, give chroma subsampling results with 4:2:2, 4:1:1 *and* 4:2:0 *schemes.*

80	88	96	12
60	8	72	68
24	52	52	8
28	20	48	12

Chroma subsampling result for 4:2:2 scheme:

80	96
60	72
24	52
28	48

[2]

Chroma subsampling result for 4:1:1 scheme:

80
60
24
28

[2]

Chroma subsampling result for 4:2:0 scheme:

(80+88+60+8)/4=59	(96+12+72+68)/4=61
(24+52+28+20)/4=31	(52+8+48+12)/4=30

[2]

6 Marks — Unseen problem

Question 2 Total Marks 27

Q3. (a) *GIF and JPEG are two commonly used image representations. What images are suitable to be represented as GIF and JPEG? Do they usually use lossless or lossy compression? Explain the reason by showing the major compression algorithm (for lossless) or the lossy steps of the algorithm (for lossy).*

Target images:

GIF : 256-colour (or 8 bit), potentially with transparency, so simple colour like graphics or drawing [1]

JPEG : continuous 24-bit true colour images [1]

Lossless or lossy:

GIF : Lossless. [1]

JPEG : Lossy. [1]

Key algorithms:

GIF : Key algorithm is LZW (lossless) [1]

JPEG : Lossy steps involve quantisation and chroma subsampling [1]

6 Marks — Bookwork

(b) *In the following situations, which (mostly)* lossless *compression algorithm is most suitable? Briefly describe the basic idea of each algorithm* in one sentence.

i. *Compression of a sequence of tokens with known, uneven probability distribution.*
Arithmetic coding: a widely used entropy coder based on range division of floating numbers. [2]

ii. *Compression of a sequence of tokens with unknown probability but with reoccurrence of patterns.*
(Lempel-Ziv-Welch) LZW coding: a compression approach to adaptively build the dictionary; only initial dictionary needs to be transmitted. [2]

iii. *Compression of a sequence of gradually changing numbers.*
Differential Pulse Code Modulation (DPCM): encode the difference between adjacent samples to reduce the dynamic range. [2]

iv. *Compression of a sequence of tokens with same tokens often appearing consecutively.*
Run-Length Encoding (RLE): map sequence to pairs of the element and the number of consecutive runs. [2]

8 Marks — Unseen problem or applied bookwork

(c) *What is the improvement of the LZW algorithm over the LZ algorithm?*

The LZW introduced the idea that only the initial dictionary needs to be transmitted to enable decoding. The decoder is able to build the rest of the table from the encoded sequence. [2]

Given the following string as input (excluding the quotes), '/THIS/IS/HIS/IS/' with the initial dictionary below, **encode** *the sequence with LZW algorithm, showing the intermediate steps.* *[11]*

Index	*Entry*
1	*/*
2	*H*
3	*I*
4	*S*
5	*T*

The steps are given as follows:

```
For: /THIS/IS/HIS/IS/

   w       k        output          index    symbol
-------------------------------------------------------
  NIL      /
   /       T          1               6        /T
   T       H          5               7        TH
   H       I          2               8        HI
   I       S          3               9        IS
   S       /          4               10       S/
   /       I          1               11       /I
   I       S
   IS      /          9               12       IS/
   /       H          1               13       /H
   H       I
   HI      S          8               14       HIS
   S       /
   S/      I          10              15       S/I
   I       S
   IS      /
   IS/     EOF        12
```

So the output will be 1 5 2 3 4 1 9 1 8 10 12

11 Marks — Unseen problem applying algorithms covered in lectures. 3 marks for keeping w, 2 marks for appropriate allocation of index, 3 marks for symbol table and 3 marks for output

Question 3 Total Marks 27

Q4. (a) *List* two *psychological phenomena that have been exploited in MEPG audio compression.*

These involve *frequency masking* and *temporal masking*. [2]

Briefly explain their meanings.

Frequency Masking : When an audio signal consists of multiple frequencies the sensitivity of the ear changes with the relative amplitude of the signals. If the frequencies are close and the amplitude of one is less than the other close frequency then the second frequency may not be heard. [1]

Temporal Masking : After the ear hears a loud sound, consisting of multiple frequencies, it takes a further short while before it can hear a quieter sound close in frequency. [1]

4 Marks: Bookwork

(b) *What is the key difference between I-frames, P-frames and B-frames in MPEG-2 video compression?*

I-Frame : Basic reference frame for each group of pictures – essentially a JPEG compressed image. [1]

P-Frame : Coded forward difference frame w.r.t. last I or P frame [1]

B-Frame : Coded backward difference frame w.r.t. last I or P frame [1]

3 Marks: Bookwork

Give the advantages and disadvantages of using B-frames.

Advantages : Improve code efficiency as most B frames use less bits; quality can be improved in the case of moving objects that reveal hidden areas; better error propagation as B frames are not used to predict future frames. [1]

Disadvantages : Frame reconstruction memory buffers within the encoder and decoder must be doubled in size to accommodate the 2 anchor frames; potentially more delays for online applications. [1]

2 Marks: Bookwork

(c) *Assume* 2×2 *macroblock is used. For the following macroblock*

#	#	#	#
#	6	4	#
#	2	2	#
#	#	#	#

the corresponding intensities in the reference frame are given as follows:

5	3	6	2
1	4	7	3
4	5	3	3
3	2	3	3

Calculate the motion vector, with complete search within ± 1 *pixel search window. List the steps to obtain the result.*

For all the 9 possibilities, compute sum of absolute difference:

-1, -1 : $|5 - 6| + |3 - 4| + |1 - 2| + |4 - 2| = 5$

-1, 0 : $|1-6|+|4-4|+|4-2|+|5-2| = 10$
-1, 1 : $|4-6|+|5-4|+|3-2|+|2-2| = 4$
0, -1 : $3-6|+|6-4|+|4-2|+|7-2| = 12$
0, 0 : $|4-6|+|7-4|+|5-2|+|3-2| = 9$
0,1 : $|5-6|+|4-3|+|2-2|+|3-2| = 3$
1, -1 : $|6-6|+|2-4|+|7-2|+|3-2| = 8$
1, 0 : $|7-6|+|4-3|+|3-2|+|3-2| = 4$
1, 1 : $|6-3|+|4-3|+|3-2|+|3-2| = 6$

[3]

So the motion vector is $(0, 1)$ with $SAD = 3$. [2]

Should the intra-frame or inter-frame coding scheme be used for this macroblock? Why?

The macroblock size $N = 2$.

$MB_{mean} = \frac{1}{4}\sum_{i=0,j=0}^{N-1} C_{i,j} = (6+4+2+2)/4 = 3.5$. [1]

$A = \sum_{i=0,j=0}^{N-1} |C_{i,j} - MB_{mean}| = |6-3.5|+|4-3.5|+|2-3.5|+|2-3.5| = 6$. [2]

$SAD - 2N = 3 - 2 \times 2 = -1$, [1]

$A > SAD - 2N$, so inter-frame coding. [1]

What is the macroblock being coded after motion compensation?

After motion compensation, the difference between the target macroblock and the best match in the reference will be used, i.e.

$$\begin{pmatrix} 6 & 4 \\ 2 & 2 \end{pmatrix} - \begin{pmatrix} 5 & 3 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}$$

[2]

12 Marks: Unseen problem

(d) *Given the following coding schemes for a group of sequential frames in MPEG-2:*

I	*B*	*B*	*P*	*B*	*B*	*B*	*P*	*I*	*B*	*B*	*B*	*P*	*B*	*P*	*P*
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*

What is the coding order of the frames?

The frame coding order is:

I	P	B	B	P	B	B	B	I	P	B	B	B	P	B	P
1	4	2	3	8	5	6	7	9	13	10	11	12	15	14	16

[6]

6 Marks: Unseen problem – first IPBBPBBB 2 marks, IPBBB 2 marks and finally PBP 2 marks

Question 4 Total Marks 27