CARDIFF UNIVERSITY
EXAMINATION PAPER

SOLUTIONS

Academic Year:	2002-2003
Examination Period:	Autumn 2002
Examination Paper Number:	CM0340
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 four pages.
There are four questions in total.

There are no appendices.

The maximum mark for the examination paper is 100% 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 THREE questions.

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

1. (a*) What is MIDI?*

Definition of MIDI: a protocol that enables computer, synthesizers, keyboards, and other musical device to communicate with each other.

2 Marks – Basic Bookwork

(b) *How is a basic MIDI message structured?*

Structure of MIDI messages:

- MIDI message includes a status byte and up to two data bytes.
- Status byte
- The most significant bit of status byte is set to 1.
- The 4 low-order bits identify which channel it belongs to (four bits produce 16 possible channels).
- The 3 remaining bits identify the message.
- The most significant bit of data byte is set to 0.

4 Marks – Basic Bookwork

(c) *A piece of music that lasts 3 minutes is to be transmitted over a network. The piece of music has 4 constituent instruments: Drums, Bass, Piano and Trumpet. The music has been recorded at CD quality (44.1 Khz, 16 bit, Stereo) and also as MIDI information, where on average the drums play 180 notes per minute, the Bass 140 notes per minute, the Piano 600 notes per minute and the trumpet 80 notes per minute.*

(i) Estimate the number of bytes required for the storage of a full performance at CD quality audio and the number of bytes for the Midi performance. You should assume that the general midi set of instruments is available for any performance of the recorded MIDI data.

CD AUDIO SIZE:

2 channels * 44,100 samples/sec * 2 bytes (16bits) * 3*60 (3 Mins) = 31,752,000 bytes = 30.3 Mb

Midi:

3 bytes per midi message

KEY THINGS TO NOTE
Need to send 4 program change (messages to set up General MIDI instruments) = 12 bytes (2 marks)
Need to send Note ON and Note OFF messages to play each note properly. (4 marks)

Then send 3 mins * 3 (midi bytes) * 2 (Note ON and OFF) * (180 + 140 + 600 + 80) = 18,000 bytes = 17.58 Kb.

8 Marks – Unseen 2 for CD AUDIO 6 for MIDI

(ii) *Estimate the time it would take to transmit each performance over a network with 64 kbps.*

CD AUDIO

Time = 31,752,000*8 (bits per second)/(64*1024) = 3,876 seconds = 1.077 Hours

MIDI

Time = 18,000*8 / (64*1024) = 2.197 seconds

2 Marks Unseen

(iii) *Briefly comment on the merits and drawbacks of each method of transmission of the performance*.

Audio: Pro: Exact reproduction of source sounds
Con: High bandwidth/long file transfer for high quality audio

MIDI: Pro: Very low bandwidth
Con: No control of quality playback of Midi sounds.

4 Marks Unseen but extended discussion on lecture material

(d) *Suppose vocals (where actual lyrics were to be sung) were required to be added to the each performance in (c) above. How might each performance be broadcast over a network?*

KEY POINT: Vocals cannot utilize MIDI

Audio: Need to overdub vocal audio on the "background" audio track
Need some audio editing package and then "mix" combined tracks for stereo audio.

Assuming no change in sample rate or bit size the new mixed track will have exactly the same file size as the previous audio track so transmission is same as in (c).

Midi: Midi alone is now no longer sufficient so how to proceed?

For best bandwidth keep backing tracks as MIDI and send Vocal track as Audio.

To achieve such a mix some specialist music production software will be needed to allow a file to be saved with synchronized Midi and Audio.

How to deliver over a network? Need to use a Multimedia standard that supports MIDI and digital audio. Quicktime files support both (as do some Macromedia Director/Flash(?) files) so save mixed MIDI audio file in this format.

The size of the file will be significantly increased due to single channel audio. If this is not compressed and assume a mono audio file file size will increase by around 15Mb. SO transmission time will increase drastically.

7 Marks Unseen

2. (a) *What is meant by the terms frequency and temporal masking of two or more audio signals? Briefly, what is cause of this masking?*

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. The range of closeness for frequency masking (*The Critical Bands*) depends on the frequencies and relative amplitudes.

Temporal Masking: After the ear hears a loud sound it takes a further short while before it can hear a quieter sound.

The cause for both types of masking is that within the human ear there are tiny hair cells that are excited by air pressure variations. Different hair cells respond to different ranges of frequencies.

Frequency Masking occurs because after excitation by one frequency further excitation by a less strong similar frequency is not possible of the same group of cells.

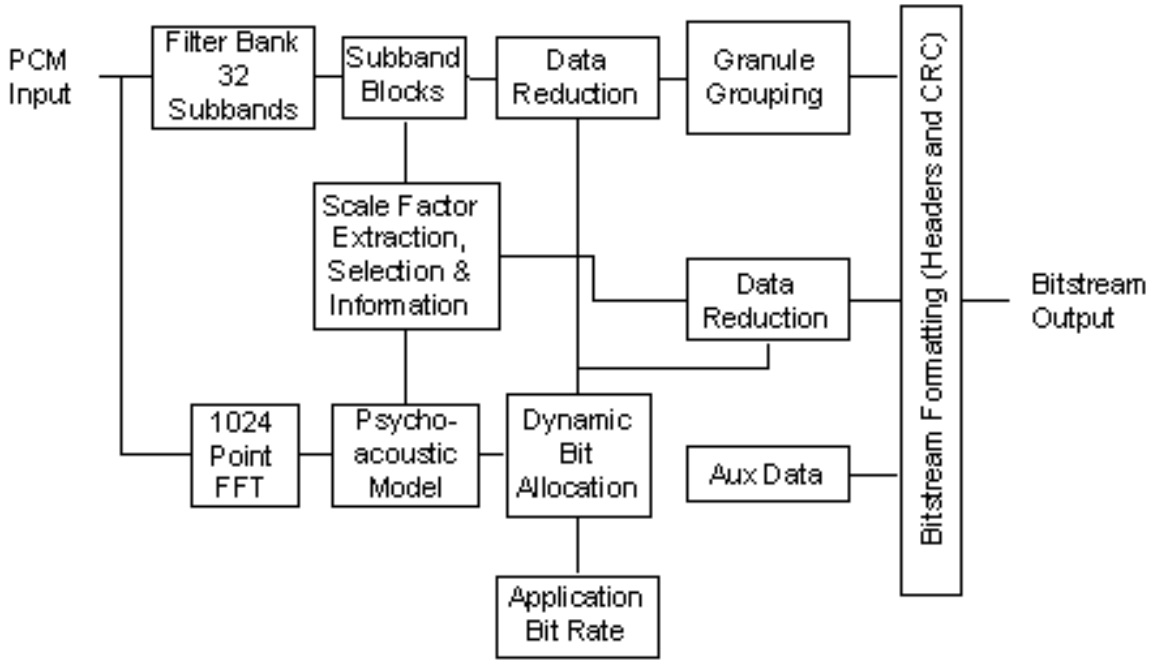
Temporal Masking occurs because the hairs take time to settle after excitation to respond again.

8 Marks – BookWork

(b) How does MPEG audio compression exploit such phenomena? Give a schematic diagram of the MPEG audio perceptual encoder.

MPEG use some perceptual coding concepts:

- Bandwidth is divided into frequency subbands using a bank of analysis filters – critical band filters.
- Each analysis filter using a scaling factor of subband max amplitudes for psychoacoustic modeling.
- FFT (DFT) used, Signal to mask ratios used for frequencies below a certain audible threshold.



8 Marks - BookWork

(c) *The critical bandwidth for average human hearing is a constant 100Hz for frequencies less than 500Hz and increases (approximately) linearly by 100 Hz for each additiional 500Hz.*

(i) Given a frequency of 300 Hz, what is the next highest (integer) frequency signal that is distinguishable by the human ear assuming the latter signal is of a substantially lower amplitude?

Trick is to realize (remember?) definition of critical band:

Critical Band: The Width of a masking area (curve) to which no signal may be heard given a first carrier signal of higher amplitude within a given frequency range as defined above.

Critical Band is 100 Hz for 300 Hz signal so if a 300 Hz Signal So range of band is 250 – 350 Hz.

So next highest Audible frequency is 351 Hz

4 Marks – Unseen

(*ii) Given a frequency of 5000 Hz, what is the next highest (Integer) frequency signal that is distinguishable by the human ear assuming the latter signal is of a substantially lower amplitude?*

5, 000 Hz critical bandwith is 10 * 100 Hz = 1000Hz

So range of of band is 4500 – 5500 Hz

So next highest audible frequency is 5501 Hz

7 Marks Unseen

3. (a) *What is the main difference between the H.261 and MPEG video compression algorithms?*

H 261 has I and P frames. Mpeg introduces additional B frame for backward interpolated prediction of frames.

2 Marks - Bookwork

(b) *MPEG has a variety of different standards, i.e. MPEG-1, MPEG-2, MPEG-4, MPEG-7 and MPEG-21. Why have such standards evolved? Give an example target application for each variant of the MPEG standard.*

Different MPEG standard have been developed for developing target domains that need different compression approaches and now formats for integration and interchange of multimedia data.

MPEG-1 was targetted at Source Input Format (SIF): Video Originally optimized to work at video resolutions of 352x240 pixels at 30 frames/sec (NTSC based) or 352x288 pixels at 25 frames/sec (PAL based) but other resolutions possible.

MPEG-2 addressed issues directly related to digital television broadcasting,

MPEG-4: Originally targeted at very low bit-rate communication (4.8 to 64 kb/sec).

MPEG-7 targetted at Multimedia Content Description Interface.

MPEG-21 targetted at Multimedia Framework: Describing and using Multimedia content in a unifed framework.

8 Marks – Bookwork

(c) *Given the following two frames of an input video show how MPEG would estimate the motion of the macroblock, highlighted in the first image, to the next frame.*

For ease of computation in your solution: you may assume that all macroblock calculations may be performed over 4x4 windows. You may also restrict your search to ± 2 pixels in horizontal and vertical direction around the original macroblock.

1	1	1	1	1	1	1	1
1	1	2	3	3	2	1	1
1	1	2	2	2	2	1	1
1	1	2	**4**	**5**	2	1	1
1	1	2	**5**	**3**	2	1	1
1	1	2	3	3	2	1	1
1	1	1	3	3	2	1	1
1	1	1	3	3	1	1	1

Frame *n*

1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1
1	1	2	1	2	2	2	2
1	1	2	1	4	3	3	2
1	1	2	1	4	3	4	3
1	1	2	1	4	4	5	4
1	1	2	1	4	5	4	5
1	1	2	1	2	4	4	4

Frame *n+1*

Basic Ideas is to search for Macroblock (MB) Within a ± 2 pixel window and work out Sum of Absolute Difference (SAD) (or Mean Absolute Error (MAE) for each window – but this is computationally more expensive) is a minimum.

Where SAD is given by:

For i = -2 to +2
For j = -2 to +2

$$SAD(i,j) = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} | C(x+k, y+l) - R(X+i+k, y+j+l) |$$

Here N = 2, (x,y) the position of the original MB, *C*, and R is the region to compute the SAD.

It is sometimes applicable for an *alpha* mask to be applied to SAD calculation to mask out certain pixels.

$$SAD(i,j) = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} | C(x+k, y+l) - R(X+i+k, y+j+l) | * (!alpha_C = 0)$$

In this case the alpha mask is not required.

So Search Area is given by dashed lines and example window SAD is given by bold dot dash line (near top right corner)

1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1
1	1	2	1	2	2	2	2
1	1	2	1	4	3	3	2
1	1	2	1	4	3	4	3
1	1	2	1	4	4	5	4
1	1	2	1	4	5	4	5
1	1	2	1	2	4	4	4

For each Window the SAD score is (take top left pixel as window origin)

	-2	-1	0	+1	+2
-2	12	12	12	11	11
-1	11	11	11	6	7
0	12	12	9	3	4
+1	11	11	9	4	5
+2	11	11	10	3	**1**

So Motion Vector is (+2, +2).

(d) *Based upon the motion estimation a decision is made on whether INTRA or INTER coding is made. What is the decision based for the coding of the macroblock motion in (c)?*

To determine INTRA/INTER MODE we do the following calculation:

$$MB_{mean} = \frac{\sum_{i=0,\, j=0}^{N-1} |C(i,j)|}{N}$$

$$A = \sum_{i=0,j=0}^{Nx,Ny} |C(i,j) - MB_{mean}| * (!alpha_C(i,j) = 0)$$

If A < (SAD – 2N) INTRA Mode is chosen.

SO for above motion

MB = 17/2 = 8.5

A = 18

So 18 is not less that (1 - 4 =) -3 so we choose INTER frame coding.

5 Marks – Unseen

4. (a) *What is the distinction between lossy and lossless data compression?*

Lossless preserves data undergoing compression, Lossy compression aims to obtain the best possible fidelity for a given bit-rate or minimizing the bit-rate to achieve a given fidelity measure but will not produce a complete facsimile of the original data.

2 Marks – Bookwork

(b) *Briefly describe the four basic types of data redundancy that data compression algorithms can apply to audio, image and video signals.*

4 Types of Compression:

- Temporal -- in 1D data, 1D signals, Audio etc.
- Spatial -- correlation between neighbouring pixels or data items
- Spectral -- correlation between colour or luminescence components. This uses the frequency domain to exploit relationships between frequency of change in data.
- Psycho-visual, psycho-acoustic -- exploit perceptual properties of the human visual system or aural system to compress data..

8 Marks – Bookwork

(c) *Encode the following steam of characters using **decimal** arithmetic coding compression:*

MEDIA

You may assume that characters occur with probabilities of
M = 0.1, E = 0.3, D = 0.3, I = 0.2 and A = 0.1.

Sort Data into largest probabilities first and make cumulative probabilities

0 - E - 0.3 - D - 0.6 – I – 0.8 – **M** - 0.9 – A – 1.0

There are only 5 Characters so there are 5 segments of width determined by the probability of the related character.

The first character to encoded is M which is in the range 0.8 – 0.9, therefore the range of the final codeword is in the range 0.8 to 0.89999…..

Each subsequent character subdivides the range 0.8 – 0.9
SO after coding M we get

0.8 - **E** - 0.83 - D - 0.86 – I – 0.88 – M - 0.89 – A – 0.9

So to code E we get range 0.8 – 0.83 SO we subdivide this range

0 - E - 0.809 - **D** - 0.818 – I – 0.824 – M - 0.827 – A – 0.83

Next range is for D so we split in the range 0.809 – 0.818

0.809 - E - 0.8117 - D - 0.8144 – **I** – 0.8162 – M - 0.8171 – A – 0.818

Next Character is I so range is from 0.8144 – 0.8162 so we get

0.8144 - E - 0.81494 - D - 0.81548 – I – 0.81584 – M - 0.81602 – **A** – 0.8162

Final Char is A which is in the range 0.81602 – 0.8162

So the completed codeword is any number in the range

0.81602 <= **codeword** < 0.8162

12 Marks – Unseen

(d) *Show how your solution to (c) would be decoded.*

Assume Codeword is 0.8161

Code can readily determine first character is M since it is in the Range 0.8 – 0.9

By expanding interval we can see that next char must be an E as it is in the range 0.8 – 0.83 and so on for all other intervals.

5 marks – Unseen