

Department of Computer Science
University College London

**Cover Sheet for Examination Paper to be sat in
March 2007**

COMPGZ05: Multimedia Systems

Time allowed 2.5 hours

Calculators are allowed

Answer any THREE questions

Checked by First Examiner:

Date:

Approved by External Examiner:

Date:

COMPGZ05: Multimedia Systems, 2007

Answer any THREE questions

Marks for each part of each question are indicated in square brackets.

Calculators are permitted

Question 1

- a. JPEG image compression involves the following six stages:
1. Convert RGB image to YUV 4:2:0 format.
 2. Divide each image up into 8x8 tiles.
 3. Convert to frequency space using a two-dimensional Discrete Cosine Transform (DCT).
 4. Quantize the frequency space, using more bits for the lower frequencies.
 5. Encode the quantized values using Run-length encoding in a zig-zag manner starting with the low frequency components.
 6. Encode the run-length encoded values using Huffman encoding.
- i. Some of these stages are lossy in that they discard information. For each stage that is lossy, state what information is discarded, and explain why JPEG can usually lose this information without adversely affecting image quality.
[8 marks]
- ii. With reference to the lossy stages above, explain why JPEG would not be a good choice of compression algorithm to use to compress the images sent by a fax machine.
[6 marks]
- iii. What is the purpose of the last two stages?
[6 marks]
- b. Summarise briefly the main ways that H.261 video encoding differs from JPEG image encoding. What options for sending a block does H.261 have that JPEG does not have?
[7 marks]
- c. MPEG 1 is very similar to H.261. Why is MPEG is better suited to playing back stored movies, whereas H.261 is better suited to video-telephony?
[6 marks]

[Total 33 Marks]

Question 2

- a. When sampling CD quality audio, the maximum audio frequency that needs to be captured is about 20KHz, as higher frequencies cannot be heard by humans. One technique that is sometimes used is called *oversampling*. With 2x oversampling, the audio data is processed as follows:
1. The analog audio is passed through an analog filter that passes sounds below 20KHz and blocks sounds above 44KHz (sounds with frequencies above 20KHz are decreased in amplitude by this filter).
 2. The filtered analog audio is then sampled 88200 times per second.
 3. Digital filter software is then used to remove frequencies in the sampled data that are above about 20 - 22KHz.
 4. Finally every second sample is discarded to reduce the sample rate to 44100 samples per second, as per a normal CD.

This (rather convoluted) procedure is used so that a cheap analog filter can be used in combination with software, rather than using a more expensive analog filter component.

- i. Suppose the analog audio contained tones at 15KHz, 30KHz, and 60KHz. What would happen to these tones in the above steps?
[3 marks]
 - ii. If the audio were directly sampled at 44100 samples per second instead of 88200 samples per second, (removing the need for step 4, but all the other steps remained the same), what would happen to the three tones?
[5 marks]
 - iii. If the audio were sampled at 44100 samples, would the digital filter perform any useful task? Justify your answer.
[5 marks]
- b. When capturing audio, if we *quantize* the data using only a few bits to represent each sample, higher frequency noise can be introduced. Explain with the aid of a simple diagram where this noise comes from.
[5 marks]
- c. In an MP3 codec, samples in the time domain are converted to the frequency domain using a Discrete Cosine Transform (DCT). Each DCT coefficient is then *requantized using fewer bits* and the quantized values are then *Huffman encoded*.

- i. With reference to the psychoacoustic model, how does MP3 decide how many bits to use for a particular DCT coefficient? Draw diagrams if these help you explain the concepts.

[10 marks]

- ii. How does Huffman Encoding reduce the bit rate in this particular case?

[5 marks]

[Total 33 Marks]

Question 3

When Skype establishes an Internet telephony audio channel from an endpoint, it can do so in three ways:

- Direct connection, using UDP.
- Indirect connection, using UDP relayed via a Supernode.
- Indirect connection, using TCP to reach a Supernode and UDP from there onwards.

a. Why might Skype provide these three modes?

[5 marks]

b. Describe the different audio problems you might expect to encounter when the first and last modes are used.

[10 marks]

c. Which mode will normally provide the best phone call?

[2 marks]

d. One of your UDP-based Skype calls happens to traverse a somewhat broken path into UCL where the traffic is being randomly load-balanced between two parallel links, so that each packet has an equal chance of taking one link or the other. One of the links has a delay of 10ms, whereas the other has a delay of 40ms. Assume Skype is operating with 20ms of audio per packet. Describe in detail what would happen to the packet stream, and what would need to happen at the Skype receiver so as to play out the audio as well as possible under the circumstances.

[9 marks]

e. Suggest two techniques that an Internet telephony application such as Skype can use to minimize the effects of packet loss. What are their relative merits?

[7 marks]

[Total 33 Marks]

Question 4

- a. MPEG 2 video can code a frame as an I frame, a P frame, or a B frame. Briefly describe the differences between these frame types. [6 marks]
- b. If MPEG 2 were used for video-conferencing, which frame types would be it be sensible to use? Justify your answer. [4 marks]
- c. How might your answer to (b) differ if the application were Internet television? [4 marks]
- d. MPEG 2 Elementary Stream video data can be wrapped up in an MPEG Programme Stream, or in an MPEG Transport Stream. Why are there two different stream formats specified in the standards? How would you best wrap MPEG video for transport over the Internet using RTP? [6 marks]
- e. If a packet of MPEG 2 video data were lost in an Internet television application, what might the effect be on the video quality of subsequent frames if the decoder attempts to decode all the remaining packets of the stream? [6 marks]
- f. Suggest two techniques that an Internet television application might use to minimize the effects of packet loss. What are their relative merits? [7 marks]

[Total 33 Marks]

Question 5

The diagram on the following page shows the main stages in the audio and video pipeline for a videoconferencing application. Assume the video is using an MPEG 1 codec at 30fps, and the audio is using G.729 audio in 20ms frames.

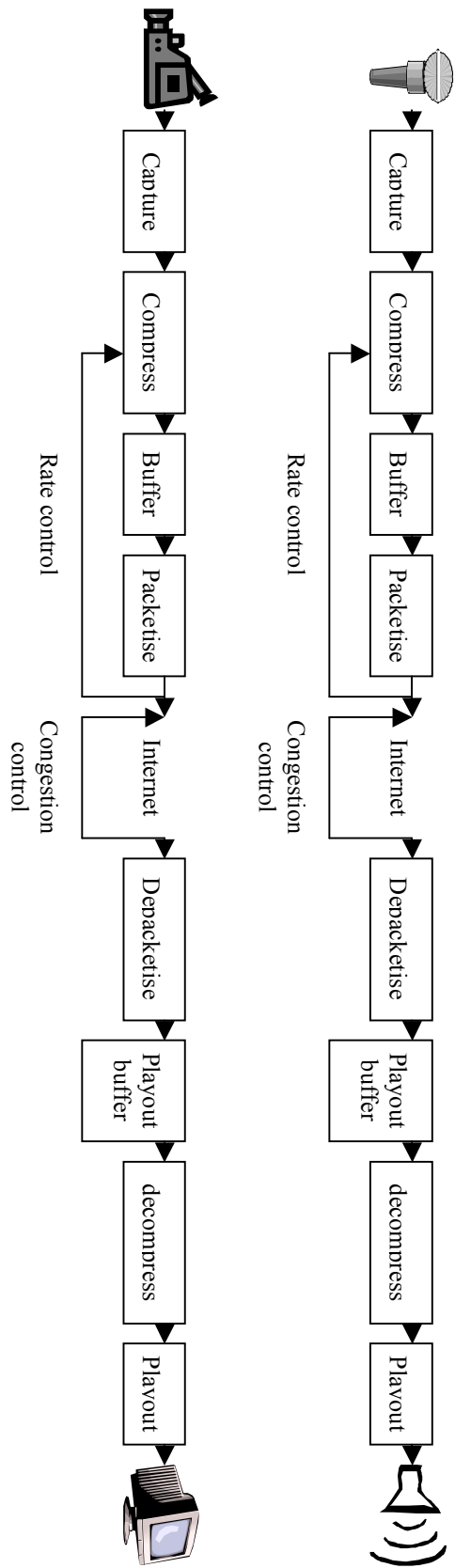
- a. What is the purpose of the buffer before packetization?
[2 marks]
- b. How might the purpose of this buffer differ between audio and video?
[5 marks]
- c. Assume the network is behaving well and providing a predictable constant throughput of 2Mb/s, with constant latency, minimal jitter, and no packet loss. Assume the congestion control is performing well, and clocking out packets at a constant rate. Furthermore, assume no inter-stream lip-synchronization is performed.
 - i. How long might data remain in the buffer before packetization in the audio pipeline?
 - ii. How long might data remain in the buffer before packetization in the video pipeline?
 - iii. How long might data remain in the playout buffer in the audio pipeline?
 - iv. How long might data remain in the playout buffer in the video pipeline?

Justify your answers and state any additional assumptions you need to make.

[12 marks]

- d. How far apart will the audio and video streams from part (c) play out?
[7 marks]
- e. How might RTP and RTCP be used to achieve lip synchronization?
[7 marks]

[Total 33 Marks]



END OF PAPER