Sample 13

1. Topic: Image Coding (24 marks)

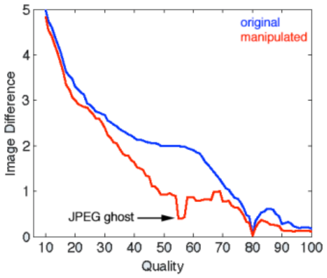
a) An image compressed using JPEG is intended to be viewed by a human observer.

i. Describe how the JPEG algorithm exploits the features of the human visual system to achieve high levels of compression while minimising visual distortion. (6 marks)

- ‘Human Visual System Features’.
- Also answers steps taken by JPEG encoder during compression.
- Colour sensitivity - Chroma subsampling is the practice of encoding images by implementing less resolution for chroma information than for luminance information, thus taking advantage of the human visual system’s lower acuity for colour differences than for luminance.
- Contrast sensitivity - DCT high frequency: For a typical image, most of the visually significant information about the image is concentrated in just a few coefficients of the DCT. Compression is achieved since the lower right values represent higher frequencies, and are often small - small enough to be neglected with little visible distortion.
- Quantisation - The compressing of image, audio or video files by removing data that does not affect the overall quality of the file i.e. lossy.
- Zig-zag scan - clusters packets of pixel information from low to high frequencies, which changes the 2D matrix into a 1D list. Arrange the coefficients in order of increasing frequency. The higher frequency coefficients are more likely to be reduced to zero after quantisation. This improves the results when using run-length encoding.
- Run Length Coding (lossless)
- Huffman Coding (lossless)

ii. Explain what is meant by the term “JPEG ghosts” and describe how they can be used to detect tampered images. (4 marks)

- This technique can determine what parts of an image underwent double compression.
- A JPEG ghost can be easily uncovered by comparing the image in question to re-saved versions of the image.
- Note that in the former case there are two dips in the graph — an expected dip at 80 (the current quality) and a second dip at 55 (the original compression quality). This second dip is a sign that the image is not an original i.e. the coefficients were previously quantized with a larger quantisation (lower quality).



b) The JPEG File Interchange Format (JFIF) is an image file format for exchanging JPEG encoded files compliant with the JPEG Interchange Format (JIF) standard.

i. Explain why JFIF is needed. (4 marks)

- JFIF enables a JPEG bitstream to be exchanged between a wide variety of platforms and applications.
- A JFIF file consists of JPEG data with a header providing information missing from the JPEG stream i.e. version number, horizontal & vertical pixel density, pixel aspect ratio and an optional thumbnail

ii. Suggest a JFIF extension that allows for transparent JPEG images. (4 marks)

- JFIF allows many extensions, for example the application markers can be utilised in order to incorporate data for transparent images. The JFIF file will contain these extensions and the JPEG data for transparent images, and if the transparent data is not needed by a particular application using the image, this extra data will simply be ignored.

(Rich's Answer)

- JFIF has many unused application markers, these can be used to extend the JPEG image
- A transparent JPEG could be made by creating and embedding a PNG mask at application marker 7
- The developer would then have to create some javascript to extract the PNG mask.
- The PNG mask could then be overlaid on the JPEG image by displaying both in a HTML5 canvas element.

iii. Outline an implementation of this extension that could be used for displaying transparent JPEG images on a web page. (6 marks)

- Mask an image. Use this mask to extract part of an embedded image at particular position. Place in alpha channel. Render alpha channel composite in JPEG image using javascript, css or HTML5 canvas blend modes.
- Produce a canvas element
- Mask the image in question. Use the mask to extract part of the embedded image at a particular position i.e. separate out the JPEG and PNG elements from the image.
- Then you want to re-draw the extracted JPEG and PNG elements in another image.
- This can be accomplished using some elements or a combination of Javascript / CSS / HTML5 canvas element.
    - ○ PRODUCE CANVAS ELEMENT
    - ○ SEPARATE OUT JPEG AND THE PNG
    - ○ USE COMPOSITING MODE IN CANVAS TO ALLOW
    - ○ DRAW JPEG
    - ○ DRAW PNG
    - ○ HOW WOULD YOU EMBED THIS AND MAIN ARCHITECTURAL ELEMNTS - WHAT ELEMENTS USED IN JS AND HTML5 TO CREATE COMPOSITE IMAGE

2. Topic: Audio and Video Coding (28 marks)

a) Describe the advantages and disadvantages of pixel-based and block-based motion representation. (4 marks)

- Blocks allow motion to be estimated between frames however the borders of moving objects rarely coincide with the borders of blocks.
- Pixels allow for a more fine grained search however at a much higher computation cost.
- bb - if block is off more visual and easy to spot, not with pb

b) MPEG-1 video compression uses I-, P- and B- pictures.

i. Explain the advantages of a mixture of picture types. (2 marks)

- The more I frames the MPEG stream has the more editable it is.
- P frames smaller than I frames so reduces the average frame size.
- B frames have the smallest frame size so reduces the average frame size.

ii. Describe a situation where video compression would not be as effective without B-pictures. (6 marks)

Notes
B-frames must be reordered so that “anchor” frames (I & P frames) are available for prediction

Answer
- B-frames allow a ‘summary’ to be made of its surrounding frames. B-frames are different to I-frames where the entire image with all its pixel values are encoded. A much smaller amount of pixels are encoded in B-frames. Thus B-frames have the smallest frame size so reduces the average frame size. This is especially important in motion video compression for example in a video clip or even a movie. If the entire movie were to be compressed using only I-frames and P-frames, its files size would be huge.
- ** B-frames are very important in reducing the file size of compressed video.

c) In telephony, the usable voice frequency band ranges from approximately 300Hz to 3400Hz. When implementing a voice codec

i. Explain why a band-pass filter would be applied to the input audio signal. (2 marks)

- This reduces the amount of data by screening out lower and higher frequencies by decomposing the signal into subbands which adds a further DCT step.

ii. Describe how the discrete cosine transform could be used to compress the speech signal. (5 marks)

- The DCT would be used to filter the speech audio signal to remove unwanted frequencies.
- The frequencies that are kept depend on the application, however in the case of speech, typically from 50Hz to 10kHz is retained. All other frequencies are blocked by the use of a band-pass filter that screens out lower and higher frequencies.

iii. Comment on the effectiveness of this approach. (3 marks)

- At the player high frequencies may reappear because of the staircase form of the signal.
- MUST FINISH WITH MILANS EXAMPLE ANSWER

d) Describe the main steps in MPEG-1 Layer 3 audio coding. (4 marks)

- MP3 coding uses a hybrid filter bank consisting of the polyphase filter bank and a Modified Discrete Cosine Transform (MDCT). The polyphase filter bank has the purpose of making Layer-3 more similar to Layer-1 & Layer-2.
- The subdivision of each polyphase frequency band into subbands increases the potential for redundancy removal.
- An estimate of the actual (time and frequency dependent) masking threshold is computed using rules known from psychoacoustics.
- The spectral components are quantized and coded with the aim of keeping the noise, which is introduced by quantizing, below the masking threshold.
- A bitstream formatter is used to assemble the bitstream, which typically consists of the quantized and coded spectral coefficients and some side information, e.g. bit allocation information.

3. Topic: Media Delivery and Presentation (28 marks)

a) ISO/IEC developed the MPEG-DASH standard allowing for dynamic adaptive streaming over HTTP.

- **Describe the main elements of an MPEG-DASH player. (6 marks)**
  - MPEG DASH streams media using small chunks of media data requested using HTTP and spliced together by the client.
  - DASH presents available content to the media player in a manifest (index) file, the Media Presentation Description (MPD), which uses an XML format.
  - Client uses HTTP to download each media segment as a sequence of files that is played back continuously.
  - DASH does not prescribe any client-specific playback functionality.

- **Describe a typical deployment architecture for MPEG-DASH. (6 marks)**
    - In a typical deployment, a DASH server provides segments in several bitrates and resolutions for example, low bitrate and high bitrate, which are encoded in video streams.
    - The video streams are then segmented into HTTP resources and a MPD file is generated for the video files, after which a URL is generated for the MPD file.
    - The DASH client may then access the MPD file based on its URL and so makes request for appropriate video files.
    - The DASH client continuously monitors and adjusts media rate based on network conditions.

b) Explain how Forward Error Correction (FEC) can be used to combat errors in wired and wireless links. Describe the conditions in which the scheme will not be effective. (4 marks)

- FEC can be used to combat errors in wired and wireless links.
- Column FEC (for Bursty Losses) where each column has a repair packet (overhead = 1/D (Depth))
- Row FEC (for Random Losses) where each row has a repair packet (overhead = 1/L (Length))
- 2D (Column x Row) FEC where overhead = (D+L) / (D x L)
- As the row and column FEC's are effectively 1 dimensional, situations in which two packets are lost side by side in the same column or same row or both would result in packet loss.

c) Explain how a decoder can conceal the loss of texture data. (4 marks)

- Spatial Concealment: the decoder can try to estimate missing pixels, blocks or groups of blocks based on the correctly received neighbouring pixels or blocks.
- Temporal Concealment: the decoder can try to estimate missing blocks or groups of blocks based on what was in the specific 'area of error' before or after the error occurred.

d) In a media streaming system

- **Explain why a client playout buffer is a key component. (2 marks)**
    - The client playout buffer is needed to account for variable network delay (jitter). As data does not always arrive at a constant rate, this buffer needs to be able to deal with times too much or too little data is being received in order to ensure smooth playback.

- **In MPEG-DASH the client controls the delivery of data. Describe an approach that an MPEG-DASH player could use to schedule media data from the server to ensure smooth playback. (6 marks)**
    - Optimise transmission rates: Here the server can plan a transmission rate so that the media can be viewed without interruption and also minimise the amount of bandwidth used.
    - To achieve this, current bandwidth consumption needs to be measured and recent behaviours need to be taken into accounts i.e. how the network has responded to similar operations recently. Used to identify whether to increase or reduce quality. The overall idea is to avoid either too much or too little data.

Sample 12

2. Topic: Video (20 marks)

b) In MPEG encoding a video sequence is divided into a Group of Pictures.

- **Describe the typical structure of a Group of Pictures, both in terms of encoding and display order. (4 marks)**

  Notes
    - A group of pictures consists of several frames. Different frame types are available according to the compression mode e.g. I-frames, P-frames, B-frames.
    - The number of frames in a GOP is application dependant.
    - MPEG Frame Types include
        - Independant reference intra frames (I-frames)
        - Predicted frames that are based on previous reference frames (P-frames)
        - Bi-directional frames that are based on previous and following frames (B-frames)

  Answer
    - A GOP in coding order always begins with an I-frame, followed by several P-frames at set distances. In the gaps are B-frames.
    - A GOP in display order must start with an I or B frame & end with an I or P frame.
    - The GOP structure is often referred to by two numbers e.g. M, N
    - M tells the distance between two anchor frames (I or P) and N tells the distance between two full images (I-frames)
    - For example if the GOP structure is IBBPBBPBBPBB
        - M=3, N=12

- **An encoder when encoding a macroblock in a P picture can decide to encode it as an INTER mode or INTRA mode macroblock. Explain why the MPEG standard allows such a choice and suggest a method that an encoder could use for deciding which macroblock mode to use. (6 marks)**
    - intra no other frames only looks at current frames
    - Both modes have different properties;

    - In INTRA mode the texture of the macroblock is coded via a DCT transformation, quantisation and entropy coding. No motion compensation is required in this mode. In this mode a frame of video is encoded as an independent image without reference to other images in the sequence.
    - INTER mode uses motion compensation. A motion compensated difference block is formed and texture coding is performed on the difference block. Regions that cannot be predicted well are coded using an image-based method. In this mode reference frames are used to predict the values

3. Topic: Audio (20 marks)

a) Describe how an audio signal is digitised. In particular, explain how the Nyquist Theorem can be used to identify the amount of data that needs to be stored. (6 marks)

- A microphone converts the sound into an electrical signal
- A filter removes very high frequencies from the signal
- ADC samples the amplitude of the analogue signal sending a stream of data to the CPU
- DAC converts a stream of numbers into a stepped analogue signal
- Smoothing filter removes the staircase shape from signal
- The Nyquist Theorem states how often we must sample in time to be able to recover the original sound.
- Nyquist: If the signal is band-limited (there is a lower limit f1 and an upper limit f2 of frequency components in the signal) then the sampling rate should be at least 2 x (f2 - f1).
- For correct sampling we must use a sampling rate equal to at least twice the maximum frequency content in the signal, this is know as the Nyquist rate (which is twice that of the Nyquist Frequency) - the maximum sampling rate to avoid aliasing. Aliasing refers to an effect that causes different signals to become indistinguishable when sampled.

b) For encoding spoken word audio (telephone) suggest (giving reasons) a suitable sampling rate, bit depth and format for storing a good quality digital audio representation. (4 marks)

- Sampling Rate: The more samples taken per second the higher the accuracy where the sampling rate depends on the application.
    - If use wrong one, important frequencies may be chopped off for example using telephony and the using it with CD’s.
    - 8kHz for telephone, 22.05kHz for radio, 44.1kHz for CD, 48kHz for a professional audio application, 96kHz for DVD audio or audio recording.
- Bit Depth: Increasing the number of bits increases the quality of the audio. 8-bit for telephone, 16-bit for CD, 20-bit for DVD audio or audio recording.
- Format: Current audio formats can support up to six channels of audio. Mono would suffice here. Stereo is good for CD’s. Surround sound for DVD’s (5 channels).

c) When performing lossy audio encoding there is a trade-off between the amount of space used and the sound quality of the result. Describe the MPEG approach to lossy audio compression. (7 marks)

- The MPEG approach to lossy compression relies on quantisation and an understanding that the human auditory system is not accurate within the width of a critical band.
- Critical bands represent the ear's resolving power for simultaneous tones.
- Psychoacoustics describe how some sounds may be masked by others under different conditions, a phenomena which can be exploited when encoding and compressing audio e.g using frequency masking, temporal masking (Answered in Summer-06)
- The audio signal is processed in discrete blocks of samples known as frames. The audio in these frames is analysed and compressed to a target number of bits using psychoacoustic modelling.

4. Topic: Delivery (20 marks)

a) Describe how a streaming media server delivers content to a client. (6 marks)

- Content is seen as a sequence of segments. Streaming allows these segments to be transmitted, segment by segment, to a user by a service provider. The streamed data can then be played on a media player, even before the entire data file is transmitted.
- Streaming allows the client to choose among stream alternatives dynamically as the network bandwidth changes.
- HTTP Live Streaming (HLS) is an example adaptive streaming protocol, which is implemented by Apple. HLS breaks a media stream into a sequence of small HTTP-based file downloads, where each download is one short chunk of an overall potentially unbounded transport system.
- A multimedia presentation is specified by a URI to a playlist file which contains a list of media URIs and information tags; these in turn specify a series of media segments. To play the stream the client obtains the playlist file and plays each segment in the playlist.

b) It is not always possible to reliably deliver a media stream via a data network such as the internet. Describe techniques (either at the server or client) that can be used to take unreliability into account. (4 marks)

- Answered in Sample-13 (FEC, client playout buffer).

c) Explain how the combination of resynchronisation markers and reversible variable length codes can increase error resilience. (4 marks)

- Without markers a single bit error in the coded bitstream prevents decoding the rest of the frame. Resynchronisation markers provide a known bit pattern interspersed throughout the bitstream.
- A reversible variable length code (VLC) ensures that code words can be decoded forward and backward. In the case of an error the data can be decoded backward from the next resynchronisation marker. More data is recovered compared to resynchronisation markers alone.

d) A media player can conceal errors in a video frame using a spatial error concealment technique. What are the advantages and disadvantages of this technique? Suggest a means of efficiently implementing this technique. (6 marks)

- Spatial Concealment: the decoder can try to estimate missing pixels, blocks or groups of blocks based on the correctly received neighbouring pixels or blocks.
- This allows errors to be corrected in frames to a certain degree. This technique requires a model of the scene and produces a blurred reconstruction.
- Repair on a pixel by pixel basis (or block by block), i.e. get average of 8 pixels surrounding missing pixel, and substitute value into missing pixel.

Summer 12

2. Topic: Video (20 marks)

a) The majority of video codecs used in practice are block-based and employ motion compensation. (4 marks)

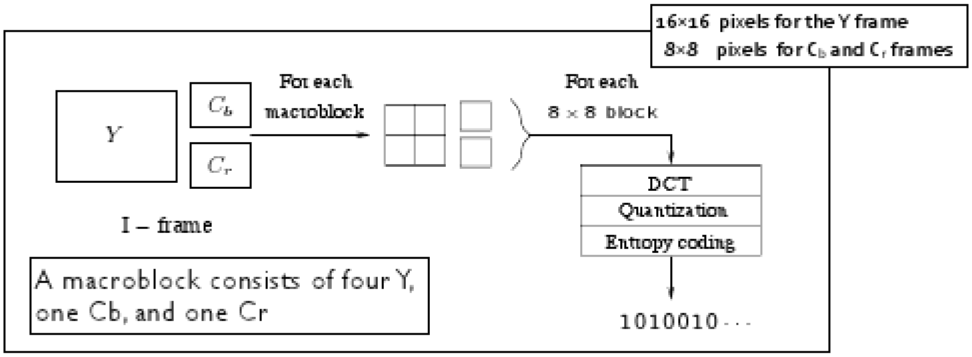
- **Explain what is meant by the term block-based encoder and why the technique is used.**
  - Block-based encoder refers to encoding carried out on frames that have been broken up into many groups of pixels which are called blocks e.g. 8x8 pixel blocks. These blocks are referred to as macroblocks where each block is equal sized, non-overlapping and rectangular. Ideally the frame dimensions is a multiple of the macroblock size. Each of these blocks can be compared with the contents of other frames for various reasons, such as to balance the effectiveness of approximating commonly seen motions, error concealment and so on.

- **Explain how motion compensation improves codec performance.**
  - Motion compensation is used to compensate inter-frame differences due to motion. The current frame to be compressed is divided into uniform, non-overlapping macroblocks. Each macroblock in the current frame is compared to the content in other frames, which is used to discover motion vectors. The motion vector detailing the position of the target macroblock's match is then encoded instead of the macroblock itself.

b) MPEG encoders have two encoding modes for individual video frames — intra (I-frames) and inter (P- and B- frames). In the former, a frame of video is encoded as an independent image without reference to other images in the sequence. In the latter, reference frames are used to predict the values. (10 marks)

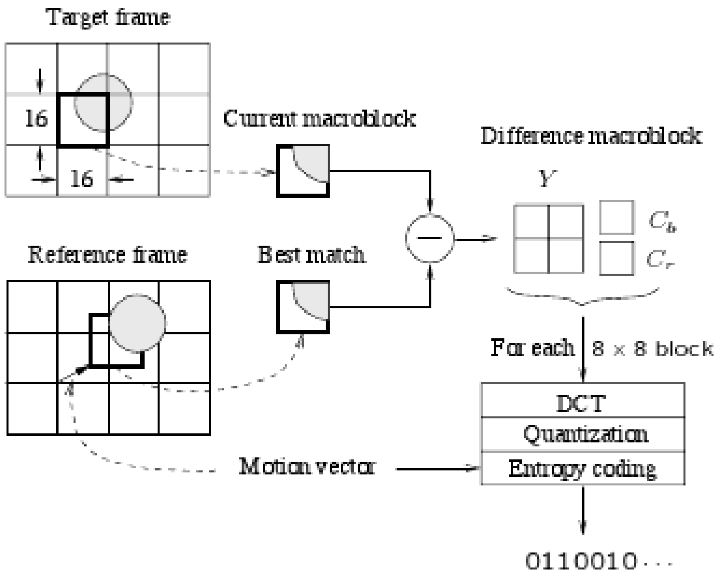
- **Describe using a diagram, the broad steps used in intra frame encoding.**
  - Decompose image into three components in RGB space & convert RGB to YCbCr
  - Divide image into several macroblocks (each macroblock has 6 blocks, 4 for Y, 1 for Cb, 1 for Cr)
  - DCT transform each block
  - Quantize each coefficient
  - Zigzag scan
  - Finally carry out Entropy Coding which involves Huffman encoding, run-length coding etc. both of which are lossless.

Compression of I-frames

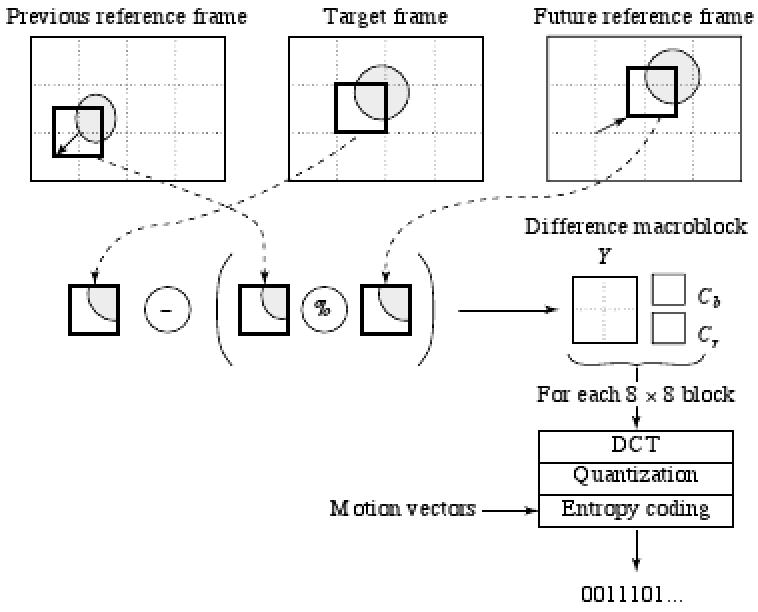

16×16 pixels for the Y frame
8×8 pixels for Cb and Cr frames
Y
Cb
Cr
For each
macroblock
For each
8 × 8 block
DCT
Quantization
Entropy coding
I – frame
A macroblock consists of four Y, one Cb, and one Cr
1010010···

- **Describe using a diagram, the key difference between encoding P- and B- frames.**
  - Each MB (macroblock) in a P-frame may have one motion vector, whereas a B-frame will have up to two motion vectors (one from forward and one from backward). If matching in both directions is successful, both corresponding MBs will be averaged before comparing against the target MB for computing the prediction error. If only one match is made (either forward or backward), its corresponding MB will be used.

P-frames



B-frames



3. Topic: Audio (20 marks)

a) 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? (2 marks)

- Think just divide frequency by two as per Nyquist.

b) Describe how Pulse Code Modulation (PCM) is used in the coding of audio data. (6 marks)

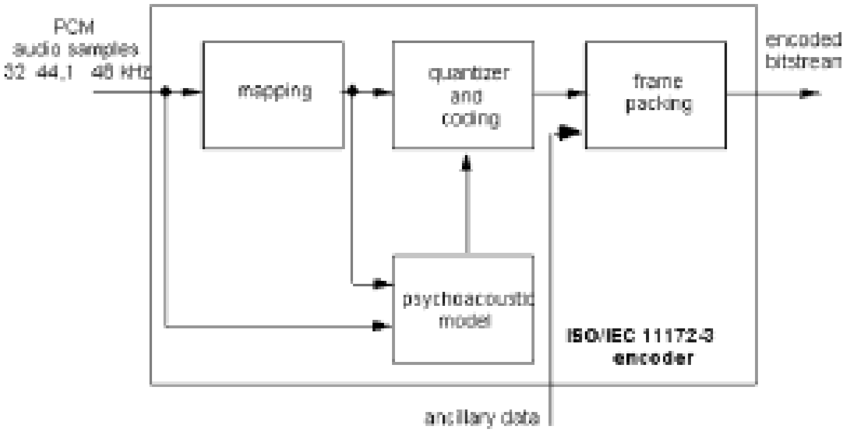
Notes

- The basic techniques for creating digital signals from analogue signals are sampling and digitisation.

Answer

- Pulse Code Modulation (PCM) is the process of
    - Sampling an analogue signal’s amplitude at fixed intervals
    - Converting the amplitude into discrete levels (quantisation)
    - Assigning digital codes to represent those levels
- PCM is also used in predictive coding, where differences are transmitted using a PCM system i.e. do not send the sample but the difference between samples.

c) Describe using a diagram the basic MPEG audio compression algorithm. (6 marks)



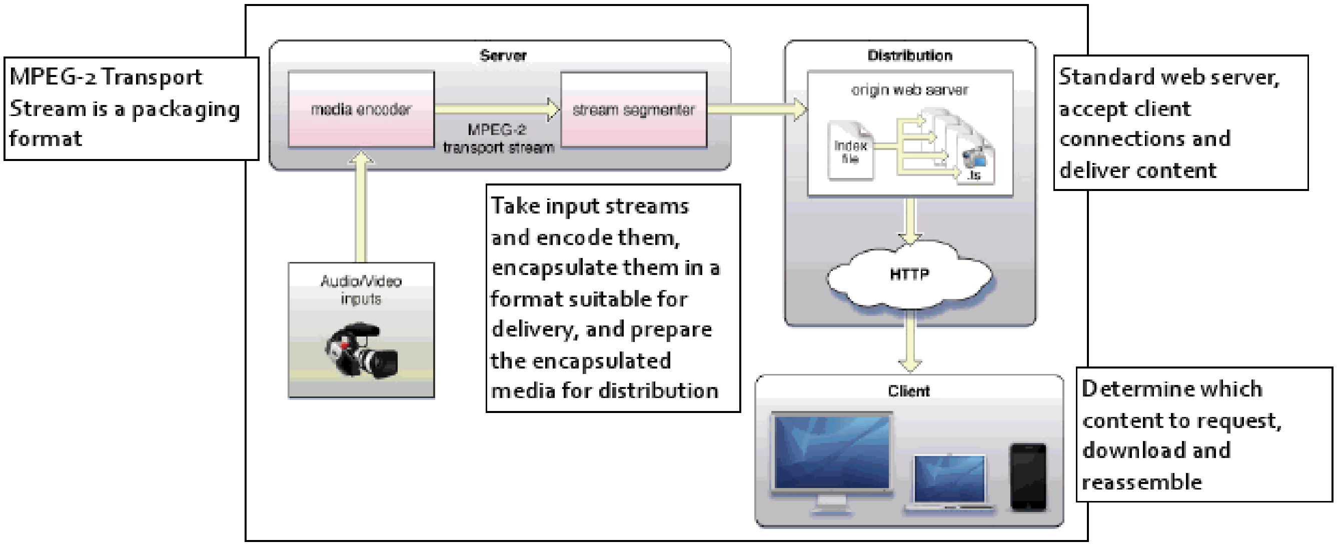
d) Describe two psychological phenomena that have been exploited in MPEG audio compression. (6 marks)

- Answered in Summer-06. (Frequency & Temporal Masking)

4. Topic: Delivery (20 marks)

a) Describe the essential elements of the HTTP Live Streaming Architecture. (6 marks)

Notes



MPEG-2 Transport Stream is a packaging format

Take input streams and encode them, encapsulate them in a format suitable for delivery, and prepare the encapsulated media for distribution

Standard web server, accept client connections and deliver content

Determine which content to request, download and reassemble

Answer

- MPEG-2 Transport Stream is a packaging format
- Audio / Video Inputs: The data to be streamed.
- Server (media encoder & stream segmenter): Take input streams and encode them, encapsulate them in a format suitable for delivery, and prepare the encapsulated media for distribution
- Distribution (origin web server & HTTP): Standard web server which accepts client connections and delivers content to the client over HTTP
- Client: Determine which content to request, download and reassemble.

d) Explain what makes loss concealment techniques feasible for digital video. (6 marks)

- Pictures are arranged as a group of frames or a group of pictures (GOP).
- Each frame may be similar to its previous or following frame. Motion of an image changes in these frames and so while pixel values may be similar throughout the frames the pixels may be in different places due to motion.
- As a result, various loss concealment techniques such as spatial & temporal concealment are made feasible.

Autumn 2012

4. Topic: Delivery (27 marks)

a) Explain how HTTP Live Streaming differs from RTSP Streaming. (9 marks)

- Live streaming covered in Sample-12. RTSP not covered.

b) It is not always possible to reliably deliver a media stream via a data network such as the internet. Describe techniques at the server and client that can be used to add resilience to the media stream. (9 marks)

- Answered in Sample-13 (FEC, client playout buffer).
- Answered in Sample-12 (Resynchronisation markers and reversible variable length codes (VLC)).

c) A media player can conceal errors in a video frame using spatial and temporal error concealment techniques. Describe how these can be implemented. (9 marks)

- Answered in Sample-12 (Spatial concealment).
- Temporal Concealment: Use temporally neighbouring areas to conceal lost regions e.g. Previous Frame Concealment (PFC)
- PFC: Use previous corresponding data to copy to current frame. Works best when there is low motion. Widely used due to simplicity of implementation.

Summer 2006

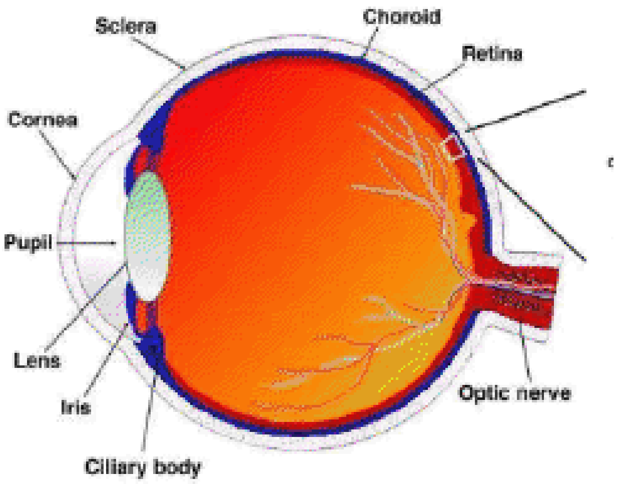
1. Topic: Fundamentals of Audio and Video Coding (20 marks)

b) Describe using an example how a digital signal is quantised. (4 marks)

- Linear and Nonlinear Quantisation
- Linear format typically stores samples as uniformly quantised values
- Non-uniform quantisation uses more finely-spaced levels where humans hear with the most acuity
- Nonlinear quantisation works by first transforming an analogue signal from the physical space into a theoretical space and then uniformly quantising the resulting theoretical space values e.g. μ-law encoding.

c) Describe using a diagram, an image formation model. (4 marks)

- The lens focuses an image onto the retina, which is upside-down and left-right reversed.
- The retina consists of an array of rods and three kinds of cones
- The rods come into play when light levels are low and produce an image in shades of grey
- The three kinds of cones are most sensitive to red (L), green (M) and blue (S)



d) Describe briefly the colour model used for digital video. (4 marks)

- YCbCr is a scaled and offset version of the YUV color space.
- YCbCr is one of two primary color spaces used to represent digital component video (the other is RGB). The difference between YCbCr and RGB is that YCbCr represents color as brightness and two color difference signals, while RGB represents color as red, green and blue. In YCbCr, the Y is the brightness (luma), Cb is blue minus luma (B-Y) and Cr is red minus luma (R-Y)

2. Topic: MPEG Standards (30 marks)

a) For MPEG-1 video describe how

ii. A good match for motion estimation is calculated. (4 marks)

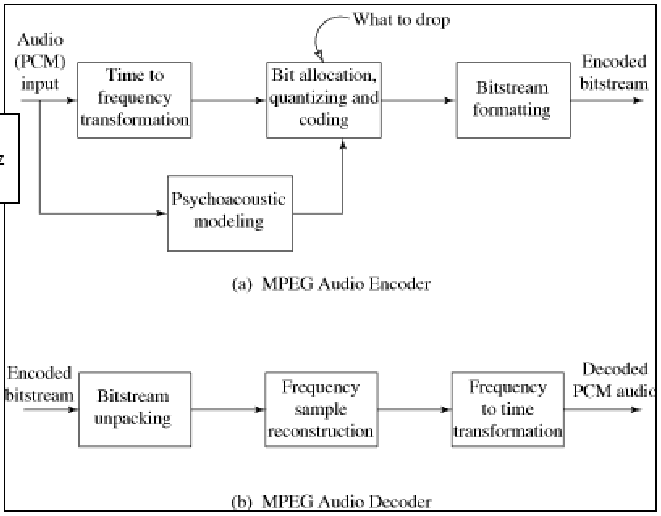
- Use a motion vector to identify where the parts of the image in the target frame were located in a reference frame
- Look for a match between the macroblock in the target frame and the most similar part of previous and/or future frame(s) known as Reference frame(s)
- The displacement of the reference block against the target macroblock is called a motion vector
- Motion estimation is a computationally intensive operation
- The amount of motion is recorded by the motion vector
    - Forward motion vectors are matches with previous frames
    - Backward motion vectors are matches with future frames

c) For MPEG audio

i. Describe how psychoacoustic techniques are used to compress audio. (6 marks)

- The threshold of hearing rises when multiple sounds come to the human ear
  - Loud sounds mask quieter sounds at nearby frequencies (Frequency Masking)
  - Loud sounds mask other sounds for a period of time (Temporal Masking)
- These phenomena can be exploited when encoding and compressing audio e.g. by means of Frequency Masking, Temporal Masking
- Frequency Masking: Lossy audio data compression methods remove some sounds which are masked. A lower tone can effectively mask (make us unable to hear) a higher tone, but a higher tone does not mask a lower tone. The greater the power in the masking tone, the wider is its influence
  - The broader the range of frequencies it can mask
  - So if two tones are widely separated in frequency then little masking occurs
- Temporal Masking: After a loud tone the ear requires time to recover. The longer a masking tone is played, the longer it takes before a test tone can be heard.

ii. Illustrate using a diagram the basic MPEG Audio encoder and decoder. (4 marks)



Sample rate 32kHz, 44.1kHz or 48kHz

Bit rate from 32kbps up to 448kbps

3. Topic: Multimedia Distribution (30 marks)

d) Describe how Quality of Service can be improved when using best-effort networks, such as the Internet, for multimedia delivery. (6 marks)

- By utilising any or all of the following
  - RSVP: signalling for resource reservations
  - Differentiated services: differential guarantees
  - Integrated Services: firm guarantees

f) Explain how damaged macroblocks can be estimated. (4 marks)

- Motion-compensated temporal interpolation: copy the corresponding macroblocks from the previous frame.
- Spatial interpolation: interpolate pixel values from pixels in adjacent correctly received macroblocks.