UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

COMP28512: Year 2 Mobile Systems

Marking Scheme – June 2013

Steve Furber & Barry Cheetham

1. **Compulsory**

Answer all of the following parts of this question. (2 marks each)

a) What is the Nyquist criterion, and how should an audio signal from a microphone be treated to ensure that this criterion is met?

The Nyquist criterion states that that the sampling frequency should be more than twice the highest frequency present in the signal. [1]

The audio signal should be passed through an anti-alias filter to ensure that no frequencies higher than half the sample frequency are present in the signal when it is sampled. [1]

b) What is a Fourier Transform, and what does it do to a time-domain signal?

A Fourier transform expresses a periodic signal as a sum of sine and cosine waves of various frequencies. [1]

It converts a time domain signal to frequency domain. [1]

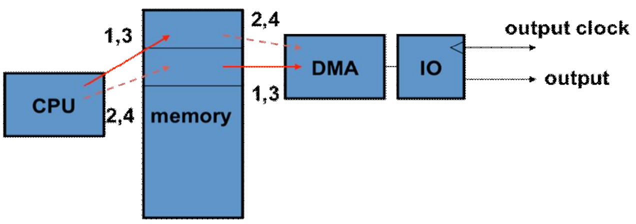
c) How much memory would be required to store 1 hour of ITU-G711 (A-Law) telephone-quality speech?

Telephone quality speech uses 8-bit samples at 8kHz sample rate [1], so 8kBytes/s, 28.8Mbytes/hr. [1]

d) What is the difference between “soft” and “hard” real-time systems?

Soft real time: a missed deadline is a loss of quality of service; acceptable if infrequent. [1]
Hard real time: a missed deadline is a system failure; unacceptable. [1]

e) Sketch the organisation in memory of a double-buffered DMA real-time data output process.



[2]

f) What is meant by interleaving and why is it beneficial when using forward error correction (FEC)?

Re-ordering a stream of bits so that adjacent bits do not remain adjacent. [1]
Advantageous because bursts of bit-errors are now spread out and become more evenly spaced and easier to correct [1]

g) What is a spectrograph (or spectrogram) and how can such a graph be conveniently displayed on a computer screen?

It is a 3D graph which shows how the short term frequency distribution of the energy of a signal varies against time. ***[1]***

The three axes are: time, frequency and spectral density with the latter plotted using colour: brighter colours for higher densities. ***[1]***

h) What does 'white' mean when describing electrical or acoustical noise and what does 'white' acoustical noise sound like?

White generally means random and spectrally flat meaning that the signal's energy is evenly spread in the frequency-domain over the available bandwidth. ***[1].***
*The sound of a waterfall, waves at the sea-side, hissing sounds when air escapes from a punctured car tyre, 'static' from a badly tuned AM or FM radio are all said to sound 'white'.****[1]***

i) What are the main properties of Huffman coding as used for psycho-acoustic music (mp3) encoding and also JPEG image encoding.

Huffman codes are variable length codes which represent the spectral samples obtained from DCT analyses of music segments or image regions. ***[1]*** *Shorter codes are used for more frequently occurring sample values and longer codes are used for less frequent sample values, and the code-words are self terminating meaning that we always know when they end.****[1]***

j) Explain the difference between waveform coding & parametric coding as applied to narrow-band speech compression.

Waveform coding techniques try to preserve exact shape of speech waveform as far as possible. Simple to implement, but cannot achieve very low bit-rates. ***[1]***
Parametric techniques represent features expected to be perceptually significant by sets of parameter. More complicated to implement, but achieve lower bit-rates. ***[1]***

2. **This question is about the use of bit-rate compression when digitising speech & music.**

a) How does 'quantisation noise' arise when digitising narrow-band speech? If a narrow-band speech coder has a 10 bit uniformly quantising analogue to digital converter and a sampling rate of 16 kHz, estimate the maximum achievable signal-to-quantisation noise ratio (SQNR) assuming the speech to be approximately sinusoidal. What is the 'dynamic range' of the encoded speech given that the SQNR must be at least 30 dB for acceptable speech quality?
State what assumptions it is reasonable to make about the nature of the quantisation noise. (6 marks)

a) Quantisation noise is noise that arises from the rounding or truncation of the true sampled values of a signal to the nearest available binary numbers when a signal is converted to digital form. ***[1]***

Maximum achievable SQNR is when signal is at its maximum amplitude without clipping or overflow (wrap-around) occurring. All 10 bits will be used. Students will know that if signal is approx sinusoidal and is uniformly quantized using N bits, SQNR = 6 x N + 1.7 dB = 61.7 dB when N=10. ***[1]***

$$\text{Dynamic range(dB)} = 10\log_{10}\left(\frac{\text{largest possible signal power representable without distortion}}{\text{smallest signal power representable with 'acceptable SQNR'.}}\right)$$

= max possible SQNR (dB) - minimum acceptable SQNR (dB) for unif quantisation

= 61.7 – 30 dB = 31.7 dB ***[1]***

Quantisation noise uniformly distrib between $\pm\Delta/2$ where Δ = step between levels. Therefore its power is $\Delta^2/12$. It also may be assumed spectrally white over the range 0 Hz to half the sampling frequency. [1]

(ii) The quantisation noise spectrum may be assumed white in the frequency range 0 to f_s / 2 Hz. [1]

In the time-domain, the quantisation error samples may be assumed random and statistically uniformly distributed between $-\Delta/2$ and $\Delta/2$ where delta is quantization step. **[1]**

b) Figure 1 shows a commonly assumed 'masking contour in quiet' for the psycho-acoustical threshold of a person's hearing.

i) Explain what this graph tells us that it is important for mp3 music coding. **(2 marks)**

ii) How could a simple experiment be performed to verify the shape of this graph for a given listener? **(2 marks)**

iii) How could simple experiments be performed to verify the effects of simultaneous (frequency) masking and temporal masking. **(4 marks)**

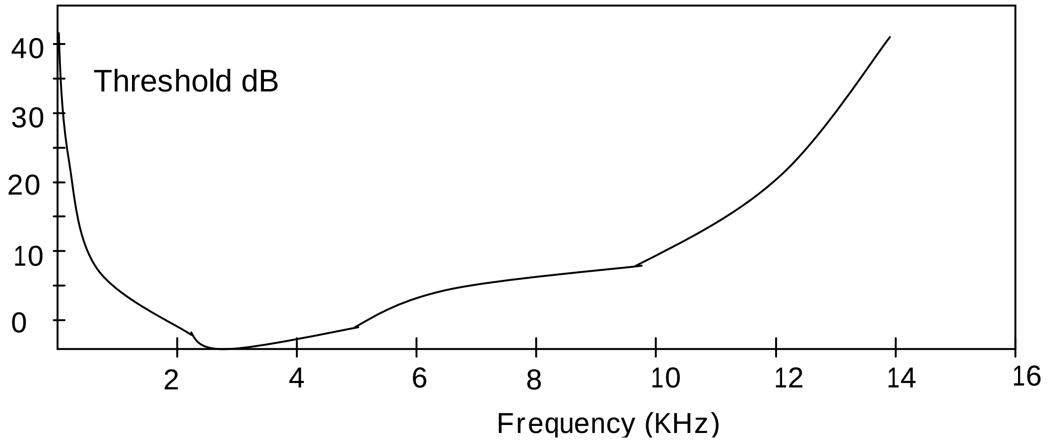


Figure 1: Masking contour 'in quiet'

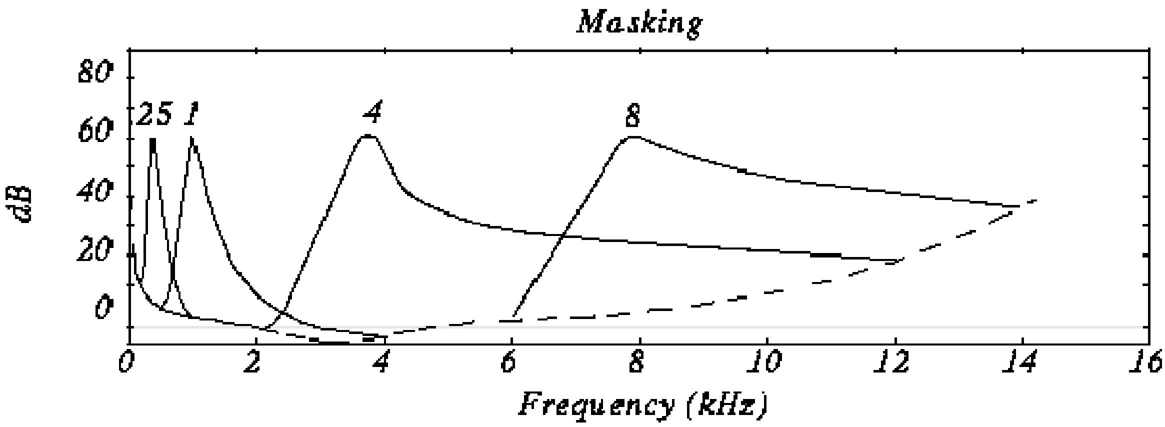
b)

i) This graph shows how the threshold of hearing varies with frequency when there are no other sounds present. **[1]**
Any sound below the threshold will not be heard and need not be encoded by an mp3 encoder. The units are in dB relative to the threshold at 2 kHz. **[1]**

ii) Generate a 2 kHz sine-wave of variable amplitude and play it out to a person over a speaker. Raise the amplitude from zero until it becomes just barely audible. Note the amplitude A(2kHz) for which this occurs. **[1]**
Repeat the test with a sine-wave of different frequency f in the range 50 Hz to 20 kHz.
Calculate threshold (dB) = $20 \log_{10}(A(f)/A(2kHz))$.
Repeat for more values of f and plot graph. **[1]**

iii) For frequency masking, play a sinusoidal 'masking tone' of frequency f (e.g. 1 kHz) at fixed amplitude. Play a 'test tone' at a different frequency (e.g. 1.1 kHz) and raise its amplitude from zero up to the value A(f) when the test tone just becomes distinguishable by the listener. Increase or decrease the frequency of the test tone and repeat the measurement. **[1]**

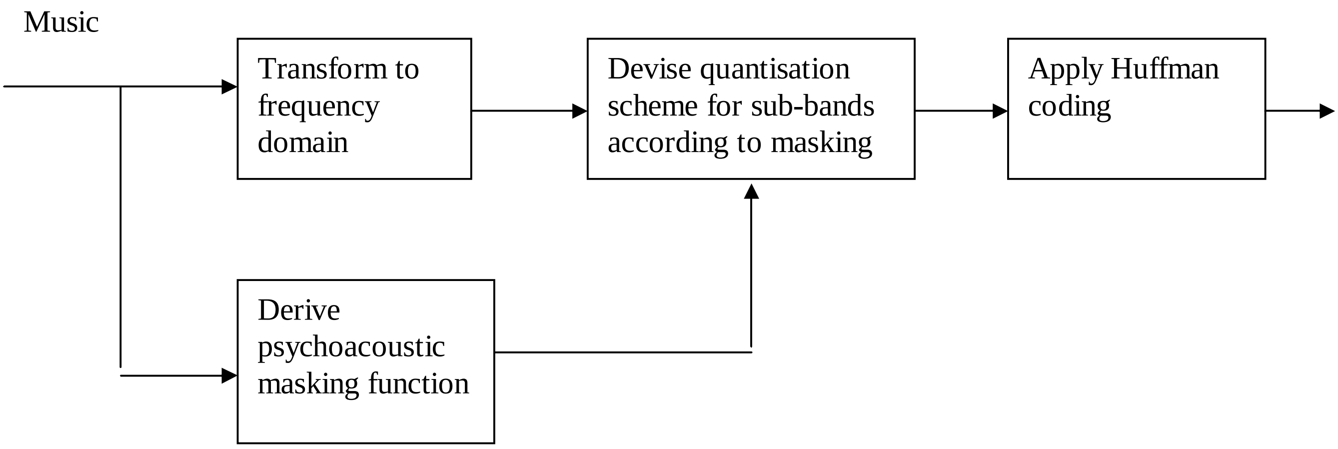
Plot the amplitude a(f) corresponding to the threshold of hearing for the test tone , in dB relative to the threshold in quiet at 2 kHz against frequency (same formula as before).
Repeat for masking tones of different frequency (e.g. 0.25, 1, 4 & 8 kHz,) always with the same fixed amplitude, to obtain graphs similar to those shown below. **[1]**



For temporal masking, play a masking tone at f Hz, & introduce a test tone at a nearby frequency with amplitude such that it is below the threshold introduced by frequency masking. Stop the masking tone, and note how much time elapses before the test tone is heard. ***[1]***
Repeat with different level of the test tone and plot how the frequency masking effect decays with time. **[1]**

c) With the aid of a block-diagram, explain how the psycho-acoustical properties of hearing mentioned above are exploited by mp3 encoders to allow high quality music to be recorded (or transmitted) at bit-rates considerably lower than what is used for compact disk recordings. **(6 marks)**

c) Block-diagram of an MP3 coder: ***[2]***



The 'Quantisation' block exploits frequency-masking by encoding accurately only the bands that will definitely be perceived. ***[1]***

For sounds in bands that will be perceived, bits are allocated according to the 'signal-to masking contour ratio' (SMR), i.e. the ratio of signal power in a particular band to the value of the masking contour at the central frequency of the band. ***[1]***

The quantisation scheme tries to make the $\Delta^2/12$ noise lower than masking threshold. ***[1]***

Also, non-uniform quantisation is used. ***[1]***

3. This question is mainly about bit-error control

a) Why is the use of forward error correction (FEC) much more important with mobile systems using wi-fi and cellular radio than with systems that use wired connections? (4 marks)

Use of FEC in cellular mobile systems increases energy efficiency & effectiveness of spatial multiplexing by frequency re-use. Transmitting at higher power is one way of making sure a signal is received with fewer errors. But high power signals carry further & cause interference over a wider range. Makes re-use of frequency bands some distance away more difficult. **[1]**
Also quickly depletes a battery powered transmitter. **[1]**
Solution is to reduce transmission power & deal with resulting increase in bit-error rate using FEC. Solves cellular frequency re-use problem & reduces power consumption. **[1]**
Also reduces any potential health risks due to the effects radiation (if there are any). **[1]**

b) A 4-bit integer 'B_3 B_2 B_1 B_0' is Hamming coded by appending three additional bits P_1, P_2 and P_3 to allow the correction of any single bit-error that may result from its radio transmission by a mobile system. Show how a suitable set of three additional bits may be derived for the given 4-bit integer and explain how the correction would be done, if necessary, at the receiver. (8 marks)

3 (b)
Make B_0 B_2 B_3 P_1 have even parity by choosing $P_1 = B_0 \oplus B_2 \oplus B_3$ *(miss out B_1)*
Make B_0 B_1 B_3 P_2 have even parity by choosing $P_2 = B_0 \oplus B_1 \oplus B_3$ *(miss out B_2)*
Make B_0 B_1 B_2 P_3 have even parity by choosing $P_3 = B_0 \oplus B_1 \oplus B_2$ *(miss out B_3)* **[2]**

At receiver, calculate 'receiver parities'
$R_1 = B_0 \oplus B_2 \oplus B_3 \oplus P_1$ *(miss out B_1)*
$R_2 = B_0 \oplus B_1 \oplus B_3 \oplus P_2$ *(miss out B_2)*
$R_3 = B_0 \oplus B_1 \oplus B_2 \oplus P_3$ *(miss out B_3)* **[2]**

If there are no bit-errors, all three receiver parities will be even, i.e. R_1 R_2 and R_3 will be 0
If just B_0 is in error, all three parities will be odd.
If just B_1 in error parities 2 and 3 will be odd
If just B_2 in error, parities 1 and 3 will be odd
If just B_3 in error parities 1 and 2 will be odd
If P_1, P_2 or P_3 in error, parities 1, 2 or 3 (respectively) will be odd, so we can assume the data is correct, assuming that only one bit-error has occurred. **[2]**

Therefore make corrections according to the following table:

R_1	R_2	R_3	Correction needed to
0	0	0	None
0	0	1	P_3
0	1	0	P_2
0	1	1	B_1
1	0	0	P_1

1	*0*	*1*	B_2
1	*1*	*0*	B_3
1	*1*	*1*	B_0

[2]

c) Explain the principle of a cyclic redundancy check (CRC). Why are CRC often used in conjunction with convolutional coders and Viterbi decoders? (4 marks)

CRC is a block code for error detection. ***[1]***
Divide (in modulo 2 arithmetic) a polynomial representing the packet bit-stream by a generator polynomial and express the remainder in binary. Do the same division at the receiver.
If we get a different remainder, a bit-error has occurred. ***[1]***
'Generator' polynomial agreed in advance & carefully chosen. Same number of check-bits always produced. Not all combinations of bit-errors are detectable by this method. Any combination that adds or subtracts multiple of generator polynomial is not detected. *[1]*
The CRC is included with the data and allows us to detect a failure of the Viterbi decoder to correct all bit-errors. **[1]**

d) According to the Shannon-Hartley Law, what is meant by the capacity C of a communication channel? Estimate the channel capacity for a mobile radio channel of band-width 4 kHz, where the reception is affected by 'additive white Gaussian noise' (AWGN) and the signal-to-noise ratio at the receiver is 60 dB. To what extent can the use of FEC increase the channel capacity? (4 marks)

Channel capacity is the maximum bit-rate achievable with arbitrarily small bit-error rate over channel affected by 'additive white Gaussian Noise'. ***[1]***
Channel capacity: $C = B \log_2(1 + S/N)$ *bits/s*
Bandwidth B Hz. S/N *= signal power / noise power ratio (not in dB),*
SNR = 10log10(S/R) is signal-to-noise ratio in dB.
$C = B \log10(1+S/N)/ \log10(2) \approx 0.332 \log10(1+S/N)$ ***[1]***
If $S/N >> 1$, $C \approx 0.332 \times B \times SNR$ *in dB.*
In this case, $C = 0.332 \times 4000 \times 60 = 80$ *kb/s* *[1]*
To no extent can FEC increase the channel capacity. *[1]*

4. This question is concerned with still and moving image compression and transmission, together with high-level mobile system design issues.

a) Describe the steps involved in compressing a still RGB image using the JPEG algorithm. (6 marks)

1. *convert RGB to luminance + chrominance; subsample chrominance [1]*
2. *use DCT to convert 8x8 blocks to frequency domain [1]*
3. *quantize DCT coefficients using quantization table [1]*
4. *replace DC component (0,0) by its difference from the previsou block [1]*
5. *zig-zag scan to read out coefficients, using run-length encoding [1]*
6. *Huffman encode the output using default or image-specific table [1]*

b) What controls the degree of compression achieved by the JPEG algorithm? (2 marks)

The quantization table controls the quality and level of compression achieved. [2]

c) Why is it inefficient to compress a moving image by simply using JPEG to compress each image in the sequence, and how is this inefficiency overcome in MPEG compression? (2 marks)

It is inefficient because consecutive images are often very similar, so there is considerable additional redundancy to exploit. [1]
MPEG uses forward and backward interpolation between images, and motion compensation. [1]

d) A museum is considering deploying a mobile audio visual system based on a PDA with a built-in digital camera, using WiFi (8O2.11) radio communications to a central server. The idea is to display a graphical icon next to each exhibit that identifies the exhibit, and to use the camera to get an image of the icon so that the server can supply relevant audio and visual information about the exhibit.

Discuss the strengths and weaknesses of the proposed mechanisms for interpreting the icons:

i) capture the raw image and send it to the server for analysis

Simple [1], but will use a lot of comms bandwidth & hence power to transmit image. [1]

ii) compress the raw image and send a JPEG file to the server for analysis

Requires more processing on the PDA [1] but less comms bandwidth than i). [1]

iii) analyze the image on the PDA and send the exhibit ID to the server. (2 marks each)

This requires the minimum comms [1] but image analysis is compute-intensive and puts a lot of demands on the PDA. Also, the PDA would require the entire icon database to be stored locally. [1]

e) It is realised that the PDAs are likely to be stolen unless there is a good security mechanism in place. A non-technical museum manager suggests leaving the camera on the PDA on so that the server can tell where it is at all times.

i) How do you explain, in non-technical terms, why this suggestion is likely to be impractical? (3 marks)

Firstly, the job of the PDA being able to understand the image from the camera to recognize where it is is a hard computer vision problem, for which there are no good solutions as yet. [2]
Secondly, continuous analysis of images from the camera would run the battery down quickly. [1]

ii) Suggest a better security mechanism. (1 mark)

All sensible answers accepted! Attach an RFID tag to the PDA and install sensors at the exits, as is done in many shops? *[1]*