THIS PAPER IS NOT TO BE REMOVED FROM THE EXAMINATION HALLS

UNIVERSITY OF LONDON **CO3325 ZA**

BSc Examination

COMPUTING AND INFORMATION SYSTEMS, CREATIVE COMPUTING, AND COMBINED DEGREE SCHEME

Data Compression

Date and Time: Wednesday 3 May 2017: 14.30–16.45

Duration: 2 hours 15 minutes

There are THREE questions in this paper. Candidates should answer all **THREE** questions. All questions carry equal marks and full marks can be obtained for complete answers to **THREE** questions. The marks for each part of a question are indicated at the end of the part in [.] brackets.

There are 75 marks available on this paper.

A hand held calculator may be used when answering questions on this paper but it must not be pre-programmed or able to display graphics text or algebraic equations. The make and type of machine must be stated clearly on the front cover of the answer book.

Question 1

(a) Discuss the absolute limit of lossless compression with the aid of an example. What is the proportion of the files (in percentage) that can be compressed by one byte? [5]

(b) Explain, with the aid of a small example, what *B picture* means and how it works in the context of video compression. [5]

(c) Explain why the Reflected Gray code is a good representation for coding the colours of greyscale images. Derive the *Reflected Gray code* for the decimal number 11. [5]

(d) i. Explain, with the aid of an example, the concept of `bitmapped images` and `vector graphics`. [6]

ii. What are the differences between bitmapped images and vector graphics in terms of *requirements* of the computer storage capacity, and the *size* of the image files? [4]

Question 2

(a) Explain the main efficiency problem of the canonical minimum-variant Huffman encoding algorithm and how the problem can be solved by maintaining two lists. [5]

(b) Following the approach of LZW algorithm, decode step by step the tokens (`1, 1, 2, 1, 3, 3, 258, 259, 257, 261, 3`). Assume that the dictionary initially contains single characters A–Z and occupies cells at 1–256 only. Demonstrate in each step the values of the main variables $x, element$, $word$, the decoding $Output$, and the changes of the dictionary. [8]

(c) Consider $S_1 = (A, B)$, the alphabet of a binary source file in which symbols A and B occur independently with the probability distribution $P_1 = (p_A, p_B)$, respectively. Let S_2 denote the four-element extended alphabet from S_1, P_2 denote the probability distribution of S_2, H_1 denote the entropy of S_1, and H_2 denote the entropy of S_2.

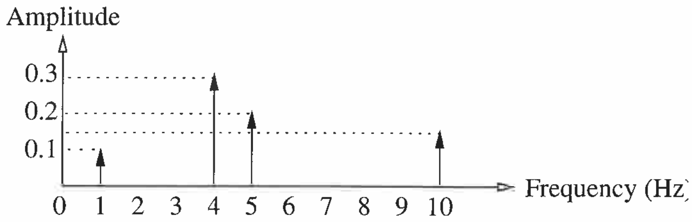
i. Write the extended alphabet S_2. [2]

ii. Suppose $S_1 = (A, B) = (0.1, 0.9)$. Compute the probability distribution P_2, entropy H_1 and entropy H_2. Show all your work. [4]

iii. Demonstrate, for any P_1, $H_2 = 2H_1$, that is, the (first-order) entropy of S_2 is double the value of the entropy of S_1. [6]
[*Hints*]: $p_{AB} = p_A \times p_B$; $log_2(p_A)^2 = 2log_2p_A$;
$log_2(p_A \times p_B) = log_2p_A + log_2p_B$; $p_A + p_B = 1$.

Question 3

(a) Explain what is meant by *sampling* in the context of data compression. Consider the frequency spectrum diagram of an analogue signal in the figure below. What is the minimum sample rate that allows the reconstruction of the signal from the samples? [4]



(b) Consider the alphabet (A, B, C, D) of a source. Discuss the possibility of finding: [6]

i. A uniquely decodable binary code in which the codeword for A is of length 2, that for B of length 1 and for both C and D of length 3.
ii. A shorter variable length prefix code than the code described in b.(i).

Provide the evidences or justify your answers.

(c) Demonstrate how adaptive Huffman encoding and decoding algorithms synchronise identical Huffman trees independently. You should use the string `CAAABB` as an example to demonstrate:

i. The encoding process [7]
ii. The decoding process. [8]

In each iteration of the process, trace the states (or values) of the `input`, `output`, `alphabet` and the `tree structure`.

END OF PAPER

THIS PAPER IS NOT TO BE REMOVED FROM THE EXAMINATION HALLS

UNIVERSITY OF LONDON **CO3325 ZB**

BSc Examination

COMPUTING AND INFORMATION SYSTEMS, CREATIVE COMPUTING, AND COMBINED DEGREE SCHEME

Data Compression

Date and Time: Wednesday 3 May 2017: 14.30–16.45

Duration: 2 hours 15 minutes

There are THREE questions in this paper. Candidates should answer all **THREE** questions. All questions carry equal marks and full marks can be obtained for complete answers to **THREE** questions. The marks for each part of a question are indicated at the end of the part in [.] brackets.

There are 75 marks available on this paper.

A hand held calculator may be used when answering questions on this paper but it must not be pre-programmed or able to display graphics text or algebraic equations. The make and type of machine must be stated clearly on the front cover of the answer book.