UNIVERSITY OF LONDON

GOLDSMITHS COLLEGE

Department of Computing

B. Sc. Examination 2017

IS52020B
Perception & Multimedia Computing

Duration: 3 hours

Date and time:

Audio track *students should answer ALL questions from part A, ONE question from part B, ALL questions from part C, and ONE question from part D.*

Graphics track *students should answer ALL questions from part A, ONE question from part B, ALL questions from part E, and ONE question from part F.*

Information for all students: *Parts A, C, and E each carry 20 marks, and parts B, C, and F each carry 30 marks. The marks for each part of a question are indicated at the end of the part in* [.] *brackets.*

This is a practical examination; you may use the computer at your desk as well as electronic calculators. Each answer requiring code or other computational material should be named according to question number, part and sub-part: for example, `Q5_b_2.pde` *for a Processing sketch in answer to part (b) sub-part (ii) of question 5. Save your answer to the exam submission folder. You are responsible for ensuring that your answers have been saved in the correct location.*

THIS PAPER MUST NOT BE REMOVED
FROM THE EXAMINATION ROOM

Part A

Term 1: Multiple Choice

Answer all questions
(All students)

Question 1

Answer all questions; more than one answer to each question might be correct, and all must be given for full credit. [20]

(a) In a normal eye, rod cells have how many pigments?

i. none
ii. one
iii. two
iv. three
v. something else

(b) Which of the following conditions can be described as red-green colourblindness?

i. protanopia
ii. deuteranopia
iii. tritanopia
iv. astigmatism
v. none of the above

(c) That blood vessels are in front of the retina in the human eye causes

i. colour blindness
ii. the blind spot
iii. blue-sky sprites
iv. receptor saturation
v. none of the above

(d) For `int i` between `0` and `int N`, which of these *Processing* expressions represents points equally-spaced around a circle?

i. (`cos(i)`, `sin(i)`)
ii. (`cos(i*TWO_PI/N)`, `sin(i*TWO_PI/N)`)
iii. (`sin(i/N*TWO_PI)`, `cos(i/N*TWO_PI)`)
iv. (`sin(i*N/TWO_PI)`, `cos(i*N/TWO_PI)`)
v. none of the above

(e) Deuteranomaly is a form of anomalous vision caused by

i. missing short-wavelength pigment
ii. mutated medium-wavelength pigment
iii. missing long-wavelength pigments
iv. missing rod pigments
v. something else

(f) Rod cells are sensitive to all wavelengths of visible light except

i. long-wavelength light
ii. medium-wavelength light
iii. short-wavelength light
iv. no exceptions
v. something else

(g) Which of the following is **not** a problem with linear interpolation between key frames?

i. discontinuities in velocity
ii. discontinuities in position
iii. lack of slow-in
iv. lack of slow-out

(h) Which of these is **not** a problem with physical modelling for animation?

i. the acceleration may be difficult to derive for a scenario
ii. the initial velocity may be unknown
iii. the relationships between objects may not be simple to model
iv. it may not be clear how to change the physical model for a different framerate

(i) The Fletcher-Munson curves describe what?

i. How perception of volume changes with frequency of a sound
ii. How perception of pitch changes with frequency of a sound
iii. How perception of amplitude changes with location of a sound
iv. How perception of pitch changes with amplitude of a sound
v. Something else

(j) Which of the following statements are **not** true about human audio perception? Select all that apply.

i. Increasing the frequency of a sine wave will increase its pitch (assuming it is audible).

ii. Increasing the amplitude of a sine wave will increase its pitch (assuming it is audible).

iii. Changing the phase of a sine wave will change its perceived location (assuming it is audible).

iv. Changing the frequency of a sine wave may change its volume (assuming it is audible).

v. Changing the shape of a waveform will change its timbre/tone colour (assuming it is audible).

(k) A 600Hz audio signal is sampled at 800Hz. At what frequency will you hear the sampled signal when it is played?

i. 200Hz

ii. 300Hz

iii. 400Hz

iv. 600Hz

v. 800Hz

(l) What is the correct order in which these processes should be applied in analog-to-digital conversion of a signal?

i. quantise, sample, low-pass filter

ii. sample, low-pass filter, quantise

iii. low-pass filter, quantise, sample

iv. sample, quantise, low-pass filter

v. low-pass filter, sample, quantise

vi. quantise, low-pass filter, sample

(m) What is the amplitude of the sine wave with the equation below?
$y = 0.6 \sin(2\pi 200t + \pi/2)$

i. 0.6
ii. 2
iii. 200
iv. $\pi/2$
v. none of the above

(n) Which of the following factors does **not** influence the reverberation of a given sound in a given space? Select all that apply.

i. The size of the space
ii. The shape of the space
iii. The materials in the space
iv. The impulse response of the given sound
v. The frequencies present in the sound

(o) Convolution can be used to achieve all **but** which one of the following effects?

i. Image motion blur
ii. Image edge detection
iii. Image sharpening
iv. Fourier transform
v. Audio reverb

(p) Which of the following statements are true about Euclidean distance? Select all that apply.

i. Euclidean distance is the only distance metric used in multimedia information retrieval
ii. When used in multimedia recommendation or information retrieval, a larger distance should indicate that two media items are more similar
iii. Euclidean distance in two dimensions can be derived using the Pythagorean theorem
iv. Euclidean distance cannot be used in more than three dimensions
v. Euclidean distance describes the "straight line distance" between two points

Part B

Term 1: Answer one of the two questions
(All students)

Question 2 Colour

(a) For each of the colour spaces below, state whether it is device-dependent or device-independent, and give an example of how that colour space is used:

i. HSB

ii. CIE XYZ

iii. sRGB

[9]

(b) Explain why none of the colour spaces in part (a) is suitable for direct use as a feature in a perceptual information retrieval system. [3]

(c) The CIE L*a*b* coordinates for a specific colour C are: {L*: 43, a*: 48, b*: 14}. Identify from the following list of colours the most perceptually similar and the most distinct-looking relative to colour C, explaining your reasoning:

P {L*: 56, a*: 50, b*: -3 }

Q {L*: 30, a*: 31, b*: 44 }

R {L*: 33, a*: 39, b*: 22 }

[5]

(d) You are making a Processing sketch in which the colour of an object changes in a perceptually smooth way. Specifically, you would like it to be Colour P at frame 0 and to smoothly change using linear interpolation so that it arrives at Colour Q at frame 10. (Assume for simplicity that you have access to a function that converts from CIE L*a*b* to RGB.)

i. Complete the equation below so that it specifies what the value for L* should be for any frame f between 0 and 10. [2]

$$L(f) = ???$$

ii. What should the value of L* be in frame 5? [2]

iii. What should the value of L* be in frame 8? [2]

(e) Suggest how you could design an information retrieval system for a collection of photographs, where the user specifies a single colour as the query. Give details of how you would compute features, what happens at query time, and what is shown to the user of your system. [7]

Question 3 Signal representations

(a) What is sampling, in the context of analog-to-digital conversion? Include a sketch illustrating the effect of sampling for audio, sensor data, or image. [4]

(b) What is quantising, in the context of analog-to-digital conversion? Include a sketch illustrating the effect of quantising for audio, sensor data, or image. [4]

(c) Using fewer quantisation bits will make an audio file smaller, but it may also reduce the perceptual quality of the file. How, precisely, is perceptual quality likely to be affected? [2]

(d) Provide a detailed critique of the following statement: If you want to make a video file smaller, all you have to do is use a lower frame rate. For example, if your original video is 30 frames per second, delete every other frame and display only 15 frames per second. [10]

(e) What is the difference between lossy and lossless compression? [2]

(f) Choose one of the following multimedia compression methods and describe how it exploits properties of multimedia signals and/or human perception to achieve compression: run-length encoding, colour indexing, predictive coding (as used in FLAC), MP3. Also explain why your chosen method is lossy or lossless. [6]

(g) Give an specific context in which a lossless compression format might be preferred. Be more specific than just saying "when quality matters." [2]

Part C

Term 2 Audio: Multiple Choice
Answer all questions
(Audio track students only)

Question 4

Answer all questions; more than one answer to each question might be correct, and all must be given for full credit. [20]

(a) A note is played on a pitched instrument, and its fundamental frequency is 300Hz. Which of the following frequencies are also likely to be present in the sound? Select all that apply.

i. 150 Hz
ii. 600 Hz
iii. 900 Hz
iv. 1000 Hz
v. None of the above

(b) At what frequency would we have to play a sine wave, for it to sound like it is the same note as the instrument above? Select the best answer.

i. 150 Hz
ii. 300 Hz
iii. 600 Hz
iv. 900 Hz
v. None of the above

(c) You compute an FFT on 1024 samples of audio. Which statement is true? Select the best answer.

i. The FFT computes 1024 bin values, each of which is a real number
ii. The FFT computes 1024 bin values, each of which is a complex number
iii. The FFT computes 512 bin values, each of which is a real number
iv. The FFT computes 2048 bin values, each of which is a complex number

(d) Which of the following statements are true about the Fourier Transform? Select all that apply.

i. The Fast Fourier Transform is fastest when applied to a number of samples which is a power of 2.
ii. When plotting magnitude spectrum output by applying the FFT to an audio signal, frequency will appear on the x-axis of the plot.
iii. When plotting magnitude spectrum output by applying the FFT to an audio signal, time will appear on the x-axis of the plot.
iv. The time complexity of the Fast Fourier Transform is $O(n \log n)$
v. The output of the FFT gives us information about both the magnitude and phase of each constituent frequency
vi. None of the above

(e) You are designing a new digital reverb plug-in using the impulse response you have measured for a reverberant cathedral. With this impulse response, you can apply reverb to any new input by doing what? Select only one answer.

i. Adding the input to the impulse response of the system

ii. Multiplying the input with the impulse response of the system

iii. Convolving the input with the impulse response of the system

iv. Delaying the input by one sample

v. None of the above

(f) Which of the following is **not** true about mel-frequency cepstral coefficients? Select just one:

i. They are widely used in speech analysis systems

ii. They are widely used in pitch tracking systems

iii. They provide a lower-dimensional representation than the FFT

iv. They can be used as an analysis feature in music recommendation systems

v. They use the mel scale to model human volume perception

(g) A set of sine waves is more likely to be heard as a single sound, rather than multiple sounds, when **all but which one** of the following statements is true:

i. The sine waves have the same phase

ii. The sine waves start at the same time

iii. The sine waves are harmonically related

iv. The sine waves come from the same location

v. The sine waves undergo the same changes in amplitude (i.e., they have the same "envelope")

(h) The diagram below shows the magnitude response and the phase response for a digital filter. What type of filter is this?

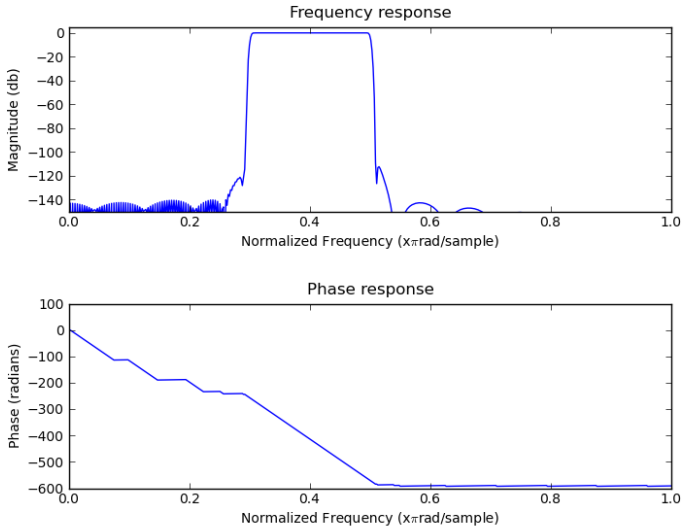
i. A low-pass filter

ii. A high-pass filter

iii. A band-pass filter

iv. A band-stop filter

v. An all-pass filter



(i) A digital filter has an impulse response $h = [0.5, 0.5]$. What type of filter is it?

i. A low-pass filter

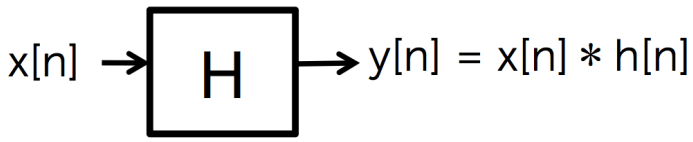
ii. A high-pass filter

iii. A band-pass filter

iv. A band-stop filter

v. An all-pass filter

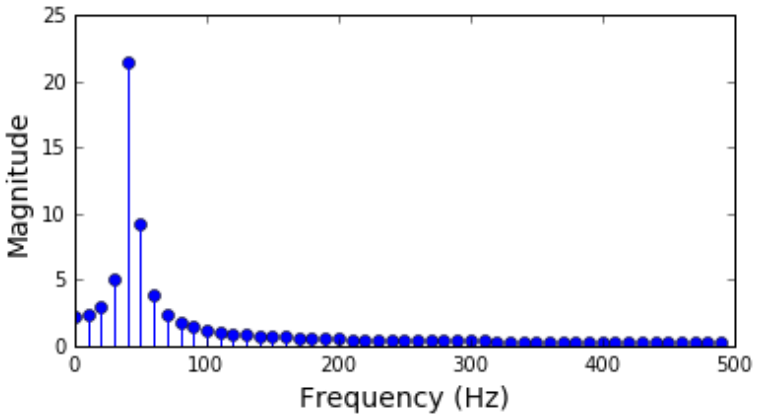
(j) What is H in the diagram below? Select one answer.



i. A signal
ii. A system
iii. An impulse
iv. An FFT
v. A duck

(k) You apply an FFT to a signal that you know is a single sine wave, which does not change in frequency over time. You get the magnitude spectrum below. Which **one** of the actions below is most likely to help you improve your ability to detect the exact frequency of this sine wave?

i. Use the STFT instead
ii. Look at the complex values output by the FFT instead of just the magnitudes
iii. Use a shorter FFT (i.e., fewer bins)
iv. Use a longer FFT (i.e., more bins)



(l) Which technique should you use to analyse the frequency content of a sound whose frequency is changing over time? Select the best answer.

i. Mel-frequency cepstral coefficients (MFCCs)
ii. Fast Fourier Transform (FFT)
iii. Short-time Fourier Transform (STFT)
iv. MP3
v. Finite Impulse Response (FIR)

(m) Which of the following are potential advantages of IIR filters over FIR filters? Select all that apply.

i. IIR filters might never stop making sound
ii. FIR filters can "explode"
iii. IIR filters are usually more computationally efficient
iv. IIR filters can be applied to both audio and sensor data, but FIR filters can only be applied to audio

(n) The FFT computes the imaginary value "$3+4j$" for a particular bin. What is the magnitude of this bin?

i. 3
ii. 4
iii. 5
iv. j
v. None of the above

(o) Which of the following statements is true? Select all that apply.

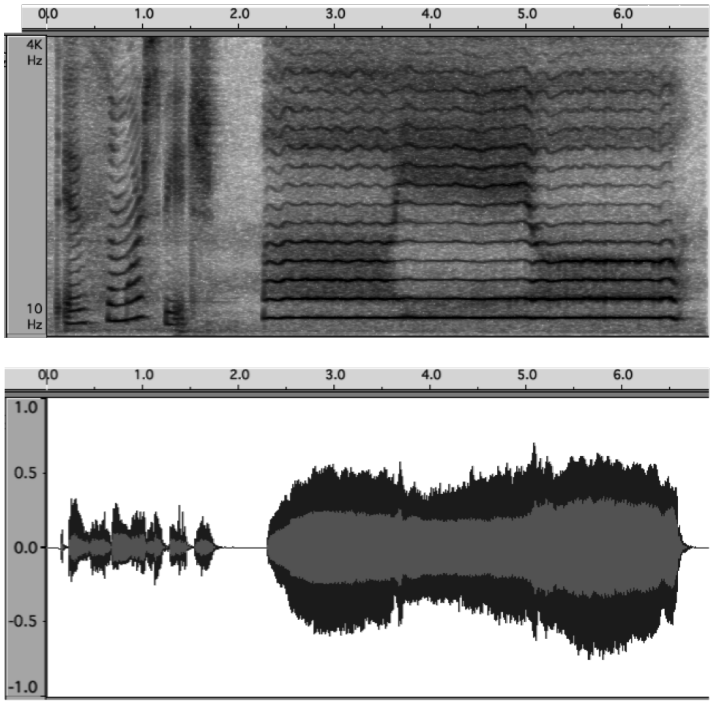
i. A multimedia fingerprint is a number
ii. A good multimedia fingerprint system will produce the same fingerprint for the same piece of media (e.g., the same photo, even if it is compressed or saved in a different file format).
iii. A good multimedia fingerprint system will produce a similar fingerprint for similar pieces of media (e.g., two songs by the same band)
iv. None of the above

Part D

Term 2 Audio: Answer one of the two questions (Audio track students only)

Question 5 Listening and Hearing

(a) The diagram below shows the spectogram (top) and waveform (bottom) for the same audio signal. Time is on the x-axis, labeled in seconds on the top. Describe as precisely as possible what you will hear if you listen to this sound. For instance, you might comment on possible source(s) of the sound, pitch (if any), volume, tone quality/timbre, the number of sound sources you hear, and how these all change over time. [7]



(b) Musical expectation

i. What factors typically cause listeners to have certain musical expectations when listening to a piece of music? [2]

ii. Give a specific example of how a piece of music might satisfy musical expectations at one moment in time, then break expectations at another moment in time. [2]

iii. Describe why a composer may want to satisfy musical expectations. That is, how does it affect the listener when their expectations are satisfied? [2]

iv. Describe why a composer may want to break musical expectations. That is, how does it affect the listener when their expectations are not satisfied? [2]

(c) Describe the phenomena of consonance and dissonance, relating our perception of these phenomena to both the physical properties of sound waves and the charac-

teristics of our inner ear. Also, discuss the relationship between consonance/dissonance perception and musical tuning. [8]

(d) You have been asked to create a new music visualiser program, in which the colour of the visualisation changes according to the current musical chord, and the size of the visualisation changes according to the volume. The input to this visualiser will be audio only: that is, you won't have access to MIDI or other data about the music; you must compute chord and loudness information directly from the audio samples.

i. Describe how you would compute information about the current musical chord. For instance, what musical feature(s) would you use, and why? [3]

ii. How well would you expect your computational approach described above to work? For instance, are there circumstances in which it is likely to fail? [2]

iii. What feature(s) would you compute from the audio in order to capture information about the current volume? Why? [2]

Question 6 Audio Signal Processing

(a) Fill in the blanks:

i. If we convolve some signal, A, with another signal, B, in the time domain, this is equivalent to _____ing the spectrum of signal A with the spectrum of signal B in the frequency domain [2]

ii. The output of a linear, time-invariant system for a given input signal is computed by convolving the input signal with _____ [2]

iii. Convolving any signal with the unit impulse will produce _____ [2]

iv. The property of _____ means that a system responds in the same manner to its inputs at all instants in time [2]

v. A system T for which the following statement is true exhibits the property of _____. [2]

$$a\mathrm{T}\{\mathrm{x}[n]\} + b\mathrm{T}\{\mathrm{y}[n]\} = \mathrm{T}\{a\mathrm{x}[n] + b\mathrm{y}[n]\}$$

(b) Your jazz fusion band has what should be an amazing recording of your most recent live show, but there are two small problems. In the middle of one of your virtuosic acoustic bass solos, your drummer's wristwatch alarm went off quite loudly, and you can hear it going "beep BEEP beep BEEP beep BEEP" in the background (it's a pitched sound, pretty high in pitch, and really annoying). Later on, in the middle of a soulful vocal solo—with no other instruments playing—the drummer dropped his sticks on the ground, which made a few "clicking" noises, also audible on the recording. Your bandmates want to fire the drummer, but he swears that these sounds can be removed from the recording using a technique he's heard of before called "filtering." Your band turns to you, the music technology expert, to resolve the dispute...

i. Do you think you'll be able to remove the wristwatch sounds using some sort of filtering approach? If so, describe in detail how you would go about it. If not, why not? [4]

ii. Do you think you'll be able to remove the stick dropping sounds using some sort of filtering approach? If so, describe in detail how you would go about it. If not, why not? [4]

(c) A waveform has the following equation:

$y = 0.5\sin(2\pi 200t) + 0.1\sin(2\pi 300t) + 0.3\sin(2\pi 600t)$

It is sampled at 2000Hz. Sketch its spectrum, with frequency on the x-axis and magnitude on the y-axis. (Do label the x-axis with specific frequencies, but don't worry about the exact values on the y-axes; just try to get the relative magnitudes approximately correct.) [5]

(d) Consider the Python signal processing example below, which begins by loading an audio waveform from a file into `signal1`.

```
signal1 = wavReadMono("w1.wav")
signal2 = hamming(2048)
tmp = signal1[0:2048]*signal2
fft_result = fft.fft(tmp)
```

i. What, exactly, is `signal2`? [1]

ii. Why, precisely, is it a good idea to multiply `signal1` by `signal2` before computing the FFT? [3]

(e) Let x be a digital signal: $x[n] = [-1, 1, 0.5]$. What is the result of convolving x with the impulse response $h = [1, -1]$? Show your work. [3]

Part E

Term 2 Graphics: Multiple Choice
Answer all questions
(Graphics track students only)

Question 7

Answer all questions; more than one answer to each question might be correct, and all must be given for full credit. [20]

(a) In what kind of coordinate system do we represent 2D points as an angle and a radius?

i. Cartesian coordinate system
ii. Polar coordinate system
iii. Homogeneous coordinate system
iv. Circular coordinate system
v. Spherical coordinate system

(b) What function does the below code perform?

$x = (r * cos(\theta))$ $y = (r * sin(\theta))$

i. Generates a sinewave and cosine wave
ii. Generates a chladni plate simulation
iii. Converts polar coordinates to cartesian coordinates
iv. Converts rectangular coordinates to polar coordinates
v. Something else

(c) In the context of working with 2D points on the complex plane, what is the below code doing?

```
x=(zx*zx)-(zy*zy);
y=2*(zx*zy);
```

i. Computing a guassian function
ii. Squaring a complex number $[zx, zy]$
iii. Getting the square root of a complex number z where $z=(x,y)$
iv. Calculating the lengths of two sides of a triangle
v. Computing a fractal tree

(d) Which of the below methods calculates the distance d between two three-dimensional points a and b?

i. $d(a,b) = pow(a * (x * x) + b * (y * y))$
ii. $d(a,b) = sqrt((x * x) + (y * y) + (z * z))$
iii. $d(a,b) = sqrt((ax - bx)^2 + (ay - by)^2 + (az - bz)^2)$
iv. $d = pow(z,x) + pow(z,y)$
v. $d = sqrt((x * x) + (y * y))$

(e) Which of the below are platonic solids? Mark all that apply.

i. Square
ii. Sphere
iii. Right-angled triangle
iv. Rhombus
v. Tetrahedron

(f) What kind of geometry does the below JavaScript code generate?

```
var points = [];
var size = 150;
var dim = 50;
var spacing = ((Math.PI * 2) / dim);

for (var i = 0; i < dim ; i++){
  var z = size * Math.cos(spacing / 2 * i);
  var s = size * Math.sin(spacing / 2 * i);
  for (var  j = 0; j < dim; j++ ) {
    var point = [Math.cos(spacing * j) * s, Math.sin(
       spacing * j) * s,z];
    points.push(point);
  }
}
```

i. face
ii. cone
iii. sphere
iv. 3D superformula
v. box

(g) In 3D graphics, what is a surface normal?

i. A vector indicating the direction that a surface should face.
ii. A point on a surface that should be used as the normal light distribution for other objects in a computer graphics scene.
iii. A vector perpendicular to the surface of a given object that can be used to calculate how the object should be lit.
iv. A function that is used to check whether or not a surface is normal.
v. A vector that represents a normal surface.

(h) What is a term for describing a piece of textured geometry used for representing the environment surrounding a 3D scene?

i. A wall
ii. A texture
iii. A skyhook
iv. A skybox
v. A space elevator

(i) In 3D graphics, what do the coordinates U and V commonly represent?

i. The contents of an image being applied to 3D geometry
ii. The coordinates of a texture from a 2D graph
iii. The texture coordinates for each vector on a surface
iv. The surface normals of a vector
v. The Unit Vector

(j) Which of the below approaches will correctly calculate a unit vector given two vectors A and B?

i. $(B + A)/\|B - A\|$
ii. $(bx + by + bz) - (ax + ay + az)/1$
iii. Get the difference V between B and A, then divide V by its magnitude.
iv. $UnitVector = (0, 0, 1) * A * B$
v. $atan(Ay/Ax)$

(k) What is a Uniform?

i. A global variable used only in C++ graphics contexts
ii. A global variable used in GLSL fragment shaders
iii. A way of providing mouse interaction in a vertex shader
iv. Three variables representing a Vector
v. A template for generating costumes for 3D characters

(l) In GLSL, what would the below code do if it was contained at the end of the main function?

```
gl_FragColor = vec4(0,1.0,0,1);
```

i. Get the coordinates of the window

ii. Set the coordinates of the current pixel

iii. Get the color of the current fragment

iv. Compute all currently used pixels

v. Make the screen go green

(m) What does the below code do?

```
vec2 p = gl_FragCoord.xy / resolution.xy
```

i. Set the resolution of the screen based on the fragment

ii. Create a new vector containing three coordinates

iii. Get a normalised value for the current fragment based on the screen resolution

iv. Makes sure that p is set to the highest possible resolution

v. Turns rendering on

(n) In a vertex shader, what is an *Attribute*?

i. A global variable

ii. A per-vertex variable

iii. A method for communicating with an external interface

iv. A parameter for a function that operates on all vertices in the same way

v. An initialisation

(o) What does the below matrix do if multiplied with a vector containing x,y,z,w values?

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

i. Rotates a 3D vector
ii. Scales a 3D vector
iii. Translates a 3D vector
iv. Nothing
v. Something else.

(p) What does the below matrix do if multiplied with a vector containing x,y,z,w values?

$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

i. Scales a 3D vector
ii. Rotates a 3D vector
iii. Translates a 3D vector
iv. Rotates and translates a 3D vector
v. Nothing

(q) What is the correct order of matrix operations when you need to increase an object's size, rotate it so it faces a new direction, then translate it to a new position?

i. scale * translation * rotation
ii. translation * scale * rotation
iii. translation * rotation * scale
iv. scale * rotation * translation
v. something else

(r) Which of the below best describes how bump mapping works?

i. By deforming geometry in the vertex shader
ii. By modifying the surface normals of 3D objects
iii. By adding noise functions to a texture in a fragment shader
iv. By scanning maps into a computer and bumping them on to a 3D object.

(s) In GLSL, what does the step function return?

i. Either a vector or a float
ii. An interpolated output
iii. 1 or 0
iv. a smooth gradient
v. a time parameter

(t) What does a tesselation shader do?

i. Creates a textured surface
ii. Generates geometric surfaces by creating new 3D objects
iii. Subdivides patches of vertex data into smaller primitives
iv. Sets the resolution of the screen
v. Divides the screen up into a grid of triangles.

Part F

Term 2 Graphics: Answer one of the two questions (Graphics track students only)

Question 8 OpenGL Shader Programming Exercise: Expressions

Use the computer to complete the following tasks, using the sample code provided as a starting point.

i. Starting from the example code provided, generate a texture with a fragment shader through the use of an expression that produces two outputs, x and y that are used to define colour values. [10]

ii. Add a time uniform. Use it to animate your function. [10]

iii. Add real-time interaction to your fragment shader using the mouse. [10]

Question 9 OpenGL Shader Programming Exercise: Distances

Use the computer to complete the following tasks, using the sample code provided as a starting point.

i. Starting from the example code, generate a circle using the fragment shader. [10]

ii. Now shade your circle so that it looks as if it is lit from above. Extra marks will be awarded for producing a 3D lighting effect. [10]

iii. Generate an image that fills the screen with circles [10]