# *Team Baseline Plan*

*Project: E-12-MJON-021 Real-time Pose Estimation*

*Supervisor: Professor Jonathan Manton*

*Team:*

1. *Matthew Yong (ID: 765 353)*
2. *Tsz Kiu Pang (ID: 965 374)*
3. *Yong Yick Wong (ID: 796 429)*

*Date: 19 April 2020*

# Table of Contents:

---

# Project Charter – E-12-MJON-021 Real-time Pose Estimation

| | |
|---|---|
| ***Author*** | Matthew Yong, Tsz Kiu Pang, Yong Yick Wong |
| ***Author ID*** | 765353; 965374; 796429 |
| ***Document Issued on*** | 23 April 2020 |
| ***Document Version No.*** | ***1.0*** |

## *Objective:*

To design and test a simple interactive sound synthesis prototype system with real-time control using gestures and upper-body movements. This prototype system is used as a "Proof-of-Concept" for a Human-Computer Interaction in the context of audio synthesis.

## *Brief Description:*

Pose estimation allows us to detect and estimate the location of human joints. In this project, we are using pose estimation to explore human-computer interaction by building a prototype of a real-time interactive system revolving audio synthesis around human gestures and movements.

The interactive system comprises three parts. The first part uses pose estimation to detect and to estimate the location of joints of our hands and upper body as observed by the computer via a webcam. The second part uses the locations detected to recognize a list of gestures such as fist and waving. The third part uses both the joint locations and gestures recognized to control parameters of audio synthesizing techniques such as frequency modulation, which would be realized in a real-time audio synthesizing environment.

This prototype system is used as a "Proof-of-Concept" for a Human-Computer Interaction in the context of real-time audio synthesis.

## Scope:

***In-Scope:***

***Must -have***
- Evaluation and final choice of open-source software for backend technologies: 2D-Pose Estimation, Gesture Recognition, and Real-time Audio Synthesizer.
- Design and testing the integration of all the technologies above either in parallel-pipelining or in a sequential flow.
- Building a library of hand and arm gestures to be recognized.
- Design of the Gesture Recognition on continuous-temporal gestures of the upper body.
- Design and testing of the mapping-and-control system to associate the gestures and the body movement to a musical parameter uniquely.
- Testing of the prototype on a set of musical parameters.

***Should-have:***
- Extend audio-synthesizing techniques.
- Extend the list of musical parameters to be controlled.
- Finger recognition from the Real-time Pose Estimation system.

***Could-have:***
- Enabling 3D Pose Estimation for more expanded control, for example hand rotation.
- Simultaneous mapping-and-control of different musical parameters in real-time.
- Real-time audio processing with audio as input.
- Generating output as Musical Instrument Digital Interfaces (MIDI) Messages.
- To expand the audio synthesizing techniques to granular synthesis and physical modeling synthesis.

***Out of Scope:***

- Offline application on local machines.
- To provide an interface for users to record and map their own gestures to the musical parameters of interest.
- Extend the prototype to different speakers for the audio synthesize to enable other features for example 3D-audio effects.
- Improving the Pose Estimation algorithm to achieve a higher control rate or frame rate.

## Stakeholders:

| | |
|---|---|
| ***Supervisor / Sponsor:*** | |
| Professor Jonathan Manton | Department of Electrical and Electronic Engineering |
| ***Expert:*** | |
| Lucas James Barbuto | Melbourne School of Engineering (Senior Systems Specialist) |
| ***External Customers / Sponsors:*** | |
| N.A. | N.A. |
| ***Team Members:*** | |
| Matthew Yong | |
| Tsz Kiu Pang | |
| Yong Yick Wong | |

## *Timing of Major Milestones:*

**For an elaborated timeline, please refer to the RACIX Matrix.*

| | |
|---|---|
| 1. Completion of the Project Charter | 24/02/2020 |
| 2. Pose Estimation System completed and tested - The system should be able to measure joint locations of interests with a PCK (Percentage of Correct Key-points) of at least 80%. | 04/05/2020 |
| 3. Gesture Recognition System completed and tested - The system should be able to recognize gestures in our list of gestures (please refer to the Scope section) with an accuracy of at least 80%. | 04/09/2020 |
| 4. Music Synthesising System completed and tested - The system should be able to generate sound based on pre-configured parameters as well as parameters controlled by gestures, with an accuracy of at least 80% | 11/09/2020 |
| 5. Final Re-design, Construction, and Testing completed | 30/09/2020 |
| 6. Endeavour Demonstration delivered | 09/10/2020 |
| 7. Presentation (Oral Exam) given | 30/10/2020 |
| 8. Final Report Submitted | 09/10/2020 |

## Success Measurement Criteria:

***1. Functional Tests***

- The system is able to detect joint locations of a pair of human hands and arms using a webcam in real-time.
- The system is able to recognize a library of recorded hands and upper body gestures in real-time.
- The recognized gestures as well as joint locations are streamed in real-time as data to an audio synthesizing environment, either Pure Data or SuperCollider.
- The recognized gestures as well as joint locations are used in real-time to control parameters in frequency modulation audio synthesizing technique, and in other techniques if time permits.
- The parameters in the audio generated should reflect the parameters controlled by the gestures.

***2. Quantitative Performance Tests***

- The joint locations of interests should have a PCK (Percentage of Correct Key-points) of at least 80%.
- Maximum latency of 7.0 seconds for a physical gesture to be recognized. This latency should be measured from the end of performing the physical gesture, to the indication of the gesture being recognized by the gesture recognition part of the system.
- Maximum latency of 8.0 seconds for a physical gesture to induce an audible change in the corresponding audio parameter. This latency should be measured from the end of performing the physical gesture, to the audible change in the corresponding parameter(s) in the audio generated by the system. This audible change could be measured by a spectrogram.
- Gestures in our library of gestures should be recognized* with an accuracy of 80%, or better.

**(please refer to the Scope section.)*

## *Critical (Important) Success Factors:*

***1. Important Inputs***

- A sufficiently powerful computing hardware to achieve real-time system:
    - Compute Unified Device Architecture (CUDA) supported Graphics Processing Unit (GPU) with at least 6.0 GB of free Random-Access-Memory (RAM) Memory.
    - or Open Computing Language (OpenCL) supported GPU of Vega-series Advanced Micro Devices Inc. (AMD) graphics card with at least 6.0 GB of free RAM Memory.
- Remote Connection Access to UniMelb server via Virtual Private Network (VPN).

***2. Important Tasks and Actions***

- To secure a Virtual Machine (VM) with computing requirements stated above from the Melbourne School of Engineering to have a shared development platform for the team.
- To set up and be able to run a real-time Pose Estimation (such as OpenPose) and Gesture Recognition Systems in the VM.
- To actively liaise with the system administrator of the VM to allocate and maintain enough computing resources continuously throughout the whole project.

## Major Risks and Treatments:

| ***Risks*** | ***Treatments*** |
|---|---|
| Potential unstable internet connection<br>- The risk could have a significant impact during the demonstration, for example at the Endeavour Exhibition. Although it is essential to have access to the virtual machine, the risk of having an unstable internet connection is relatively small compared to other risks identified. | - Use our phones as personal hotspots in the case of an unstable internet connection during the demonstration.<br>- Pre-record our demonstration |
| No redundancies in the infrastructure should it fail<br>- Since our application depends heavily on the virtual machine there would be a great impact on our project if the virtual machine fails or access is denied during the course of the project. | - Secure an alternative platform such as Google Cloud<br>- Use our personal PC to run the application, which would require an alternative library of gestures that do not require locating exactly the fingers' joints to be recognized. |
| Gestures not accurately recognized<br>- Occlusion and noisy image from webcam would reduce the accuracy of gesture recognition | - Improve the pose estimation algorithm for better accuracy in hand pose estimation in general<br>- Use our own recorded hand poses as training data to the pose estimation algorithm for better accuracy in locating joints of our own hand poses<br>- Choose gestures that are easier (with less occlusion) to be recognized when building our library of gestures (as stated in the Scope section) |
| Unfamiliarity with gesture recognition<br>- Since our team is not particularly familiar with gesture recognition, there could be unforeseeable scope creep associated with it, which might require extra time to be dealt with. | - Familiarise ourselves with gesture recognition at the early stage of our project, so that if scope creep does appear, it could be dealt with earlier and the project timeline could be modified earlier if necessary. |

| | |
|---|---|
| Overall Probability of Technical Success | 85%<br>Pose Estimation and Gesture Recognition is well known in theory, but not completely familiar to the Project Team - Challenging Development. |

## Cost – Budgetary Estimate:

| Item/Category | Number | Unit Cost | Subtotal |
|---|---|---|---|
| Hours (students) | 200 hours x 3 =600 | $100 per hour | $60,000.00 |
| Hours (supervisor - Professor Jonathan) | 25 hours | $200 per hour | $5,000.00 |
| Hours (expert - Mr. Lucas) | 8 hours | $200 per hour | $1,600.00 |
| | | ***Total Cost:*** | $66,600.00 |

## Signatures:

| Role | Name | Signature | Date |
|---|---|---|---|
| ***Supervisor*** | Professor Jonathan Manton | | |
| ***Team Member*** | Matthew Yong | Matthew | 19 April 2020 |
| ***Team Member*** | Tsz Kiu Pang | Tsz Kiu | 19 April 2020 |
| ***Team Member*** | Yong Yick Wong | Yick | 19 April 2020 |

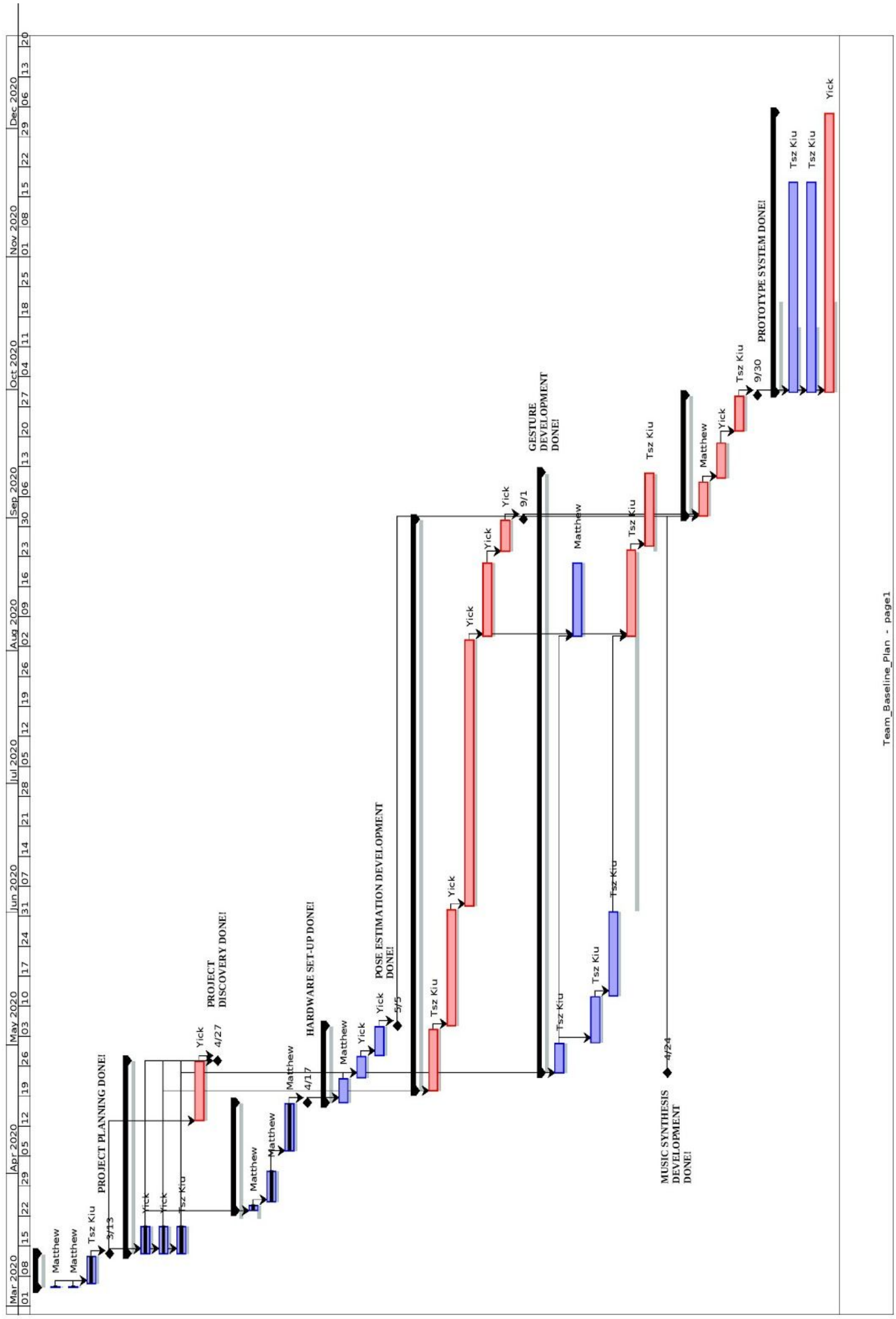
## References:

| Document | Version No. | Date |
|---|---|---|
| Brief Description of Project (Real-time Pose Estimation) | 1.0 | 02 March 2020 |
| Skeleton Charter by Bob Warfield | 1.3 | 01 March 2019 |
| ENGR90037 Lecture Material: Capstone Project Management Lecture 01, Week 03 | 1.0 | 17 April 2020 |
| Research Paper: "Neural Networks for Mapping Hand Gestures to Sound Synthesis Parameters" Author: "Paul Modler" Source: http://www.music.mcgill.ca/~mwanderley/Trends/Trends_in_Gestural_Control_of_Music/DOS/P.Mod.pdf | 1.0 | September 1998 |
| Research Paper: "Real-time 2D Hand Pose Estimation with Simultaneous Region Localisation" Source: https://www.yangangwang.com/papers/WANG-SRH-2019-11.pdf | 1.0 | September 2019 |

# Gantt Chart

Note*:
- The Critical Paths are indicated in RED.
- Tasks have been excluded in the Chart. Please refer to the RACIX Matrix.
- Milestones are indicated by diamond symbols.

**With reference to the Gantt Chart.*


PROJECT PLANNING DONE!
PROJECT DISCOVERY DONE!
HARDWARE SET-UP DONE!
POSE ESTIMATION DEVELOPMENT DONE!
GESTURE DEVELOPMENT DONE!
MUSIC SYNTHESIS DEVELOPMENT DONE!
PROTOTYPE SYSTEM DONE!
Team_Baseline_Plan - page1

# RACIX Matrix

Note*:
- "A' indicates who is accountable for the Task.
- "R" indicates who is responsible for the Task.
- "C" indicates whether the supervisor needs to be consulted.
- "X" indicates signed-off.
- Resource-Names are equivalent to "A", but it's the Software's Feature.
- Stakeholders:
    - Supervisor:  Professor Jonathan Manton *(Jonathan)*
    - IT Senior System Specialist: Lucas James Barbuto *(Lucas)*

**With reference to the RACIX MATRIX.*

| No. | Tasks | Duration (days) | Start | Finish | Predecessors | Matthew | Tsz Kiu | Yick | Jonathan | Lucas (IT) |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | **Project Planning** | 8 | Thu, 5/3/20 | Fri, 13/3/20 | - | - | - | - | - | - |
| 3 | Plan Project Management and Collaboration Tools | 1 | Thu, 5/3/20 | Fri, 6/3/20 | - | A | R | R | - | - |
| 4 | Setting up Meeting Logistics | 1 | Thu, 5/3/20 | Fri, 6/3/20 | - | A | R | R | C | - |
| 5 | Plan Project Proposal | 7 | Fri, 6/3/20 | Fri, 13/3/20 | 3,4 | R | A | R | I | - |
| 6 | *Project Planning Done!* | 0 | Fri, 13/3/20 | Fri, 13/3/20 | 5 | - | - | - | X | - |
| 7 | **Project Discovery** | 42 | Fri, 13/3/20 | Fri, 24/4/20 | - | - | - | - | - | - |
| 8 | Choosing Pose Esitmation Algorithm | 7 | Fri, 13/3/20 | Fri, 20/3/20 | 6 | R | R | A | - | - |
| 9 | Researching on Gesture Recognition Techniques | 7 | Fri, 13/3/20 | Fri, 20/3/20 | 6 | R | R | A | - | - |
| 10 | Discovering Music Synthesis Software | 7 | Fri, 13/3/20 | Fri, 20/3/20 | 6 | R | A | R | - | - |
| 11 | Documenting Team Baseline Plan (Assignment 1) | 11 | Mon, 13/4/20 | Fri, 24/4/20 | 6 | R | R | A | X | - |
| 12 | *Project Discovery Done!* | 0 | Fri, 24/4/20 | Fri, 24/4/20 | 8,9,10,11 | - | - | - | I | - |
| 13 | **Hardware Set-Up** | 20 | Sat, 21/3/20 | Fri, 10/4/20 | - | | - | - | - | - |
| 14 | Requesting for Virtual Machine from MSE-IT Support | 4 | Sat, 21/3/20 | Wed, 25/3/20 | 8 | A | R | R | I | R |
| 15 | Installing software environment for Pose Estimation to run on | 8 | Wed, 25/3/20 | Thu, 2/4/20 | 14 | A | R | R | - | C |
| 16 | Setting up Audio and Video live streaming | 4 | Mon, 6/4/20 | Fri, 10/4/20 | 15 | A | R | R | - | C |
| 17 | *Successfully run pose estimation program!* | 0 | Fri, 10/4/20 | Fri, 10/4/20 | 16 | - | - | - | I | - |
| 18 | **Pose Estimation System Development** | 24 | Fri, 10/4/20 | Mon, 4/5/20 | - | - | - | - | - | - |
| 19 | Develop communication system between Pose Estimation and Gesture Recognition Program | 5 | Fri, 10/4/20 | Wed, 15/4/20 | 17 | A | R | R | - | - |
| 20 | Reduce pose estimation GPU consumption down to GPU Specifications | 3 | Wed, 15/4/20 | Sat, 18/4/20 | 19 | R | R | A | - | C |
| 21 | Run tests and check accuracy of output keypoints | 16 | Sat, 18/4/20 | Mon, 4/5/20 | 20 | R | R | A | - | - |
| 22 | *Pose Estimation Algorithm Done!* | 0 | Mon, 4/5/20 | Mon, 4/5/20 | 21 | - | - | - | I | - |
| 23 | **Gesture Recognition System Development** | 137 | Mon, 20/4/20 | Fri, 4/9/20 | - | - | - | - | - | - |
| 24 | Defining a list of workable gestures to map using keypoints | 11 | Mon, 20/4/20 | Fri, 1/5/20 | 9 | R | A | R | - | - |
| 25 | Finding and creating training datasets of human body gestures | 21 | Fri, 1/5/20 | Fri, 22/5/20 | 24 | R | R | A | - | - |
| 26 | Creating a software system to recognize human gestures with Artificial Intelligence | 77 | Fri, 22/5/20 | Fri, 7/8/20 | 25 | R | R | A | - | - |
| 27 | Running tests to ensure both pose and gesture programs meet GPU specifciations | 14 | Fri, 7/8/20 | Fri, 21/8/20 | 26 | R | R | A | C | C |
| 28 | Running tests datasets on Gesture Program to observe accuracy | 14 | Fri, 21/8/20 | Fri, 4/9/20 | 27 | R | R | A | C | - |
| 29 | *Gesture Recognition System Done!* | 0 | Fri, 4/9/20 | Fri, 4/9/20 | 28 | - | - | - | I | - |
| 30 | **Music Synthesis System Development** | 76 | Fri, 24/4/20 | Fri, 11/9/20 | - | - | - | - | - | - |
| 31 | Choosing a real-time audio synthesising software | 4 | Fri, 24/4/20 | Tue, 28/4/20 | 10 | R | A | R | - | - |
| 32 | Developing a communication system between Gesture and Music program | 14 | Fri, 7/8/20 | Fri, 21/8/20 | 26,31 | A | R | R | - | - |
| 33 | Researching audio synthesising techniques of interest | 7 | Mon, 20/4/20 | Mon, 27/4/20 | - | R | A | R | - | - |
| 34 | Defining Musical Parameters to Control | 16 | Mon, 20/4/20 | Wed, 6/5/20 | 33 | R | A | R | - | - |
| 35 | Creating a Gesture to Musical Parameter Control System | 20 | Tue, 4/8/20 | Mon, 24/8/20 | 34 | R | A | R | - | - |
| 36 | Running tests to ensure the generated audio reflects the parameters controlled by gestures | 15 | Thu, 27/8/20 | Fri, 11/9/20 | 35 | R | A | R | C | - |
| 37 | *Music Synthesis System Done!* | 0 | Fri, 11/9/20 | Fri, 11/9/20 | 36 | - | - | - | I | - |
| 38 | **Integrating into Single Prototype System** | 21 | Fri, 4/9/20 | Fri, 9/10/20 | - | - | - | - | - | - |
| 39 | Testing Live streaming output of each subsystem | 7 | Fri, 4/9/20 | Fri, 11/9/20 | 22,29,37 | A | R | R | - | - |
| 40 | Complete Tests based on standard from a defined rubric | 7 | Fri, 18/9/20 | Fri, 25/9/20 | 39 | R | R | A | C | - |
| 41 | Documenting test results | 7 | Fri, 2/10/20 | Fri, 9/10/20 | 40 | R | A | R | I | - |
| 42 | *Prototype System Done!* | 0 | Fri, 9/10/20 | Fri, 9/10/20 | 41 | - | - | - | I | - |
| 43 | **Project Presentation** | 21 | Fri, 9/10/20 | Fri, 30/10/20 | - | - | - | - | - | - |
| 44 | Endeavour Exhitbition | 7 | Fri, 9/10/20 | Fri, 16/10/20 | 42 | R | A | R | X | - |
| 45 | Oral Presentation | 14 | Fri, 30/10/20 | Fri, 13/11/20 | 42 | R | A | R | X | - |
| 46 | Final Report | 21 | Fri, 9/10/20 | Fri, 30/10/20 | 42 | R | R | A | X | - |

# *END OF DOCUMENT*