|  |  |
| --- | --- |
| A picture of a winding road and trees  Speech radar  16/02/19 | **Student name:** shah ali  **degree title:** COMputer science  **supervisor name:** Jamie ward  **registration number:** 33455846 |

**Contents page**

1. **Introduction………………………………………………………………………………………………………………………2**
2. **Aims and objectives………………………………………………………………………………………………………….2**
   1. Aim and objective 1: successful implementation………………………………………………………2
   2. Aim and objective 2: accent detection & appropriate dataset………………………………….2
3. **Methods……………………………………………………………………………………………………………………………3**
   1. Aim and objective 1 step-by-step……………………………………………………………………………..3
   2. Aim and objective 2 step-by-step……………………………………………………………………………..3
4. **Project plan………………………………………………………………………………………………………………………4**
   1. Design & research…………………………………………………………………………………………………….4
   2. Implementation, preliminary report & review………………………………………………………….4
   3. Implementation (continued)…………………………………………………………………………………….5
   4. Testing……………………………………………………………………………………………………………………..5
   5. Final report………………………………………………………………………………………………………………6
5. **Progress to date………………………………………………………………………………………………………………..7**
   1. Design phase………………………………………………………………………………………………………..
   2. Research phase
   3. Implementation phase
   4. Summary
6. **Planned work**
7. **Appendices**
8. **Reference list**

**1. Introduction**

Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum Lorem ipsum

**2. Aims and objectives**

**a. Aim and objective 1: successful implementation**

The situation of losing sight of your phone and not knowing its location is very common amongst all of us. Some people will try to ring their phone and follow the sound to uncover it. Some people will locate it via GPS. And some people will look everywhere for it tirelessly. The aim of my project is to enable users to search for their phone in a more modern and efficient way. Speech recognition is something that's improving on a daily basis through inventions such as Amazon Alexa and Google Hub. I aim to implement this amazing deep learning task to the best of my abilities and test it thoroughly, so its detection rate is at an appropriate level and is something users can rely on. The way I will achieve this aim is by using TensorFlow and Android Studio interchangeably. TensorFlow is an open-source library that is great in tackling most deep learning problems and is easy to use once you get use to it. Android Studio is going to be our IDE where most of the implementation is going to take place. The great thing about Android Studio is that it enables users to design their app efficiently and as a bonus it works well with TensorFlow.

**b. Aim and objective 2: accent detection and appropriate dataset**

At the beginning of the project, I aimed to create a speech recognition software that could predict majority of words in the English dictionary and detect different accents at the same time. Based on recent research, I realised that I should aim to only detect a small set of words of around 30 – 60, as I would need to search for millions of audio clips and data to cover the full dictionary. Given the time I have for this project it will simply not be possible for me to gather it all. Thus, I intend to create a speech recognition that can detect different accents, but with a dataset that is appropriate for the time I have.

**3. Methods**

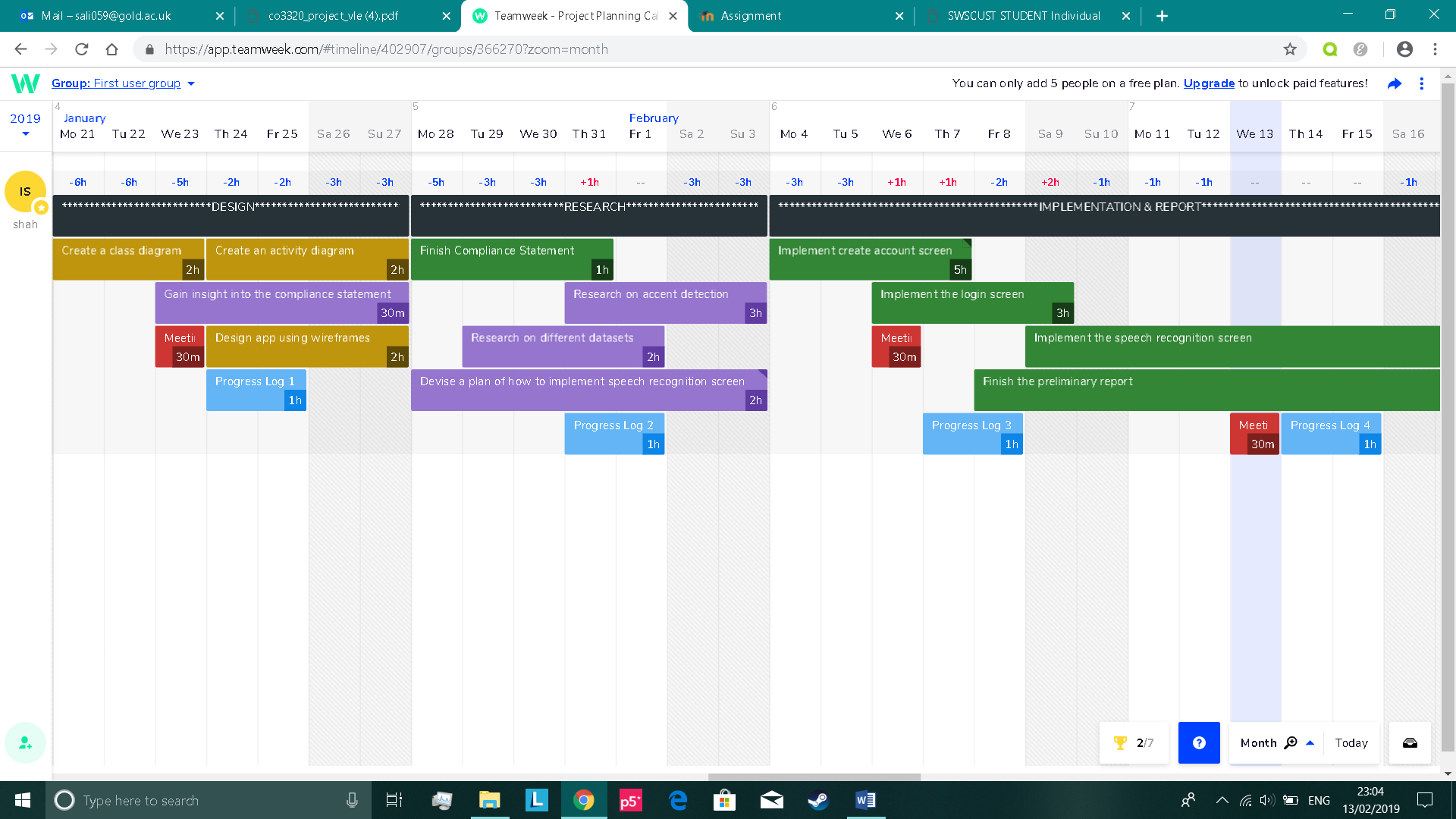
**a. Aim and objective 1 step-by-step**

* Create a deep neural network in TensorFlow, which will be trained, validated and tested. The validation set will assist me in understanding whether the model is good with unseen data. The test set will help in measuring the predictive power of the model
* Then, export the TensorFlow model as a protobuf file and place it in the assets folder within Android Studio. Next, compile TensorFlow into Android Studio via the gradle. This will allow me to use the TensorFlowInferenceInterface class, which will be needed to feed in inputs and to get the prediction results
* Once the above step is complete, get the model to load successfully
* Once the model has been loaded, insert a button that records audio in the app. Store the audio in a short array, convert it into a floating-point array and then pass the results into the defined ‘recognise’ function, which should predict the utterance from the user and choose the correct label and its probability
* Finally, I will get the app to run in the background of the users phone and have it detect voices at the same time. This is so it can identify if user utters the code word they registered (e.g. frog.) If it detects it, then the phone will start vibrating and ringing loudly. The link in the references section show exactly how I could achieve this in Android Studio
* The Github link mentioned in references assisted me in creating the plan above

**b. Aim and objective 2 step-by-step**

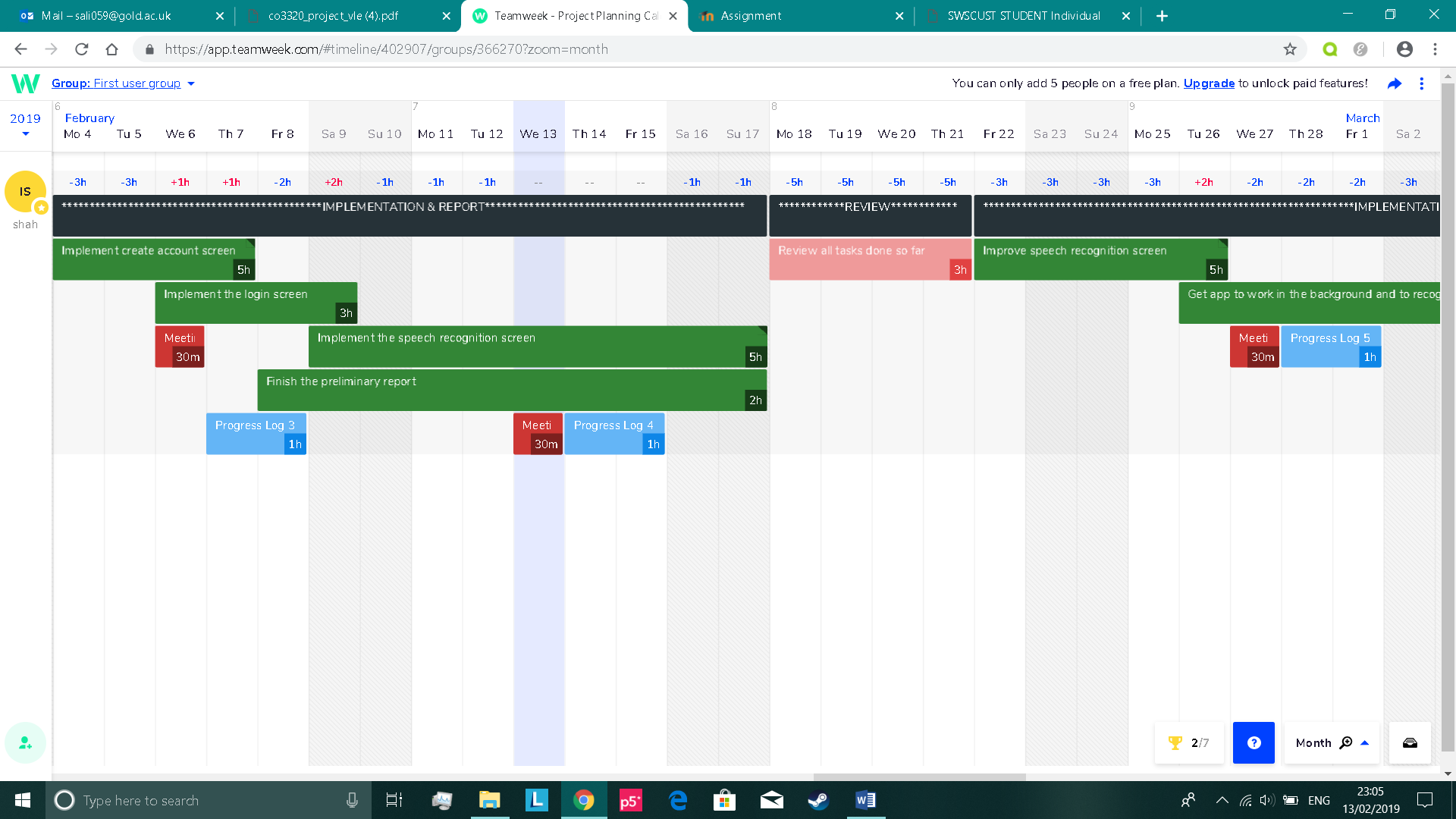
* I will use the Kaggle speech command dataset for my TensorFlow model, in order to classify one second audio clips of around 30 words. There are thousands of audio clips for each word uttered in different accents. This will be useful in enhancing the detection rate

**4. Project plan**

**a. Design & Research**

**Sub-tasks:**

1. Use pen and paper to start off the plan
2. Then develop on the plan using diagrams and flow-charts
3. Gain insight through research in whether the plan is feasible

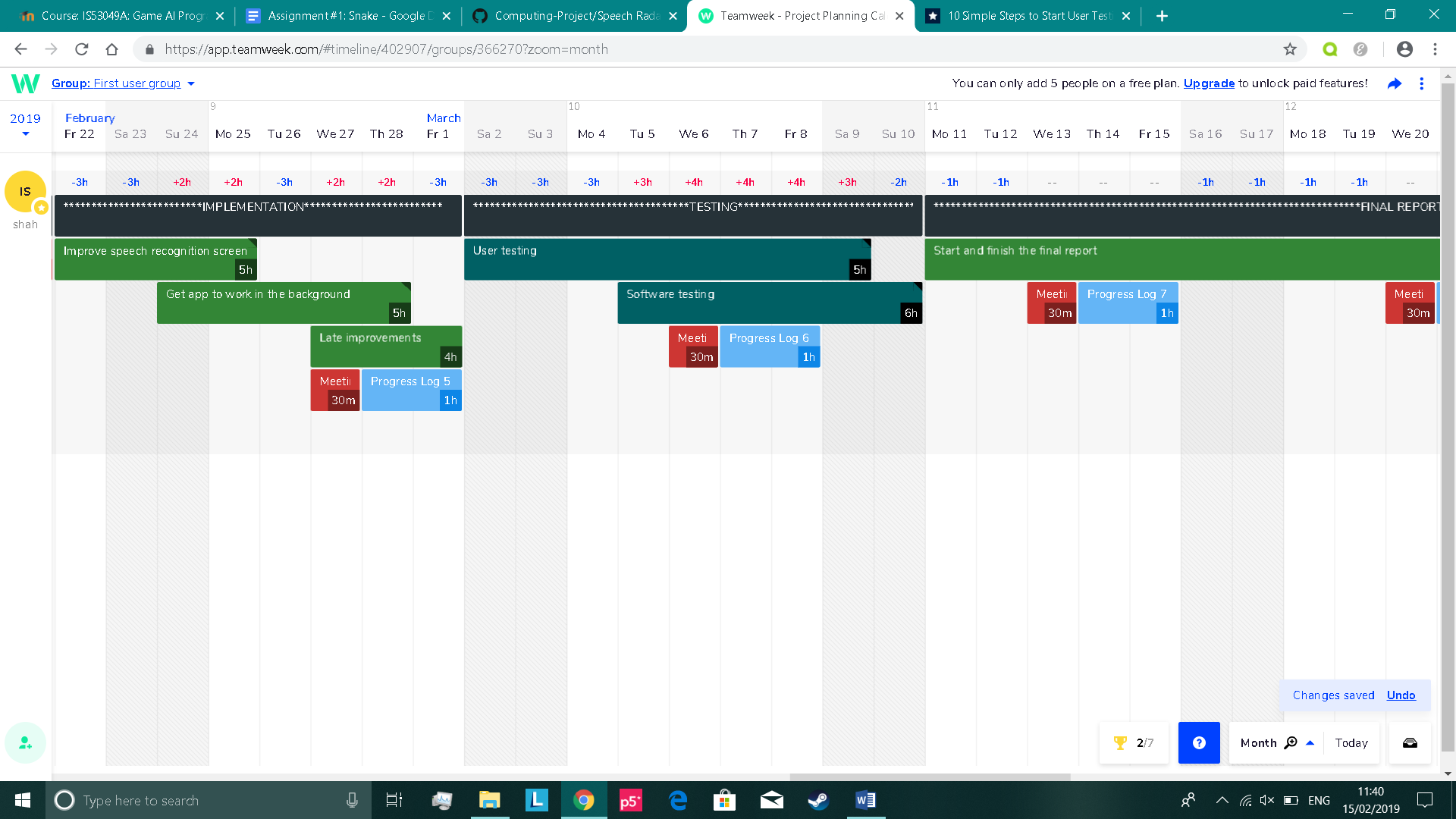
**b. Implementation, Preliminary Report & Review**

**Sub-tasks:**

1. Validate each input field
2. Insert the data into the Firebase Database once user clicks on the create account button
3. Send an email verification to the inputted email address

**Sub-tasks:**

1. Create the tensorflow model first
2. Export the model as a protobuf (.pb) file once finished
3. Move the file to assets folder in Android Studio
4. Successfully load the model
5. Create a button that'll record audio
6. Convert the audio into a floating point array
7. Feed the array into the predefined recognise function
8. Test the application and see if it makes predictions accurately

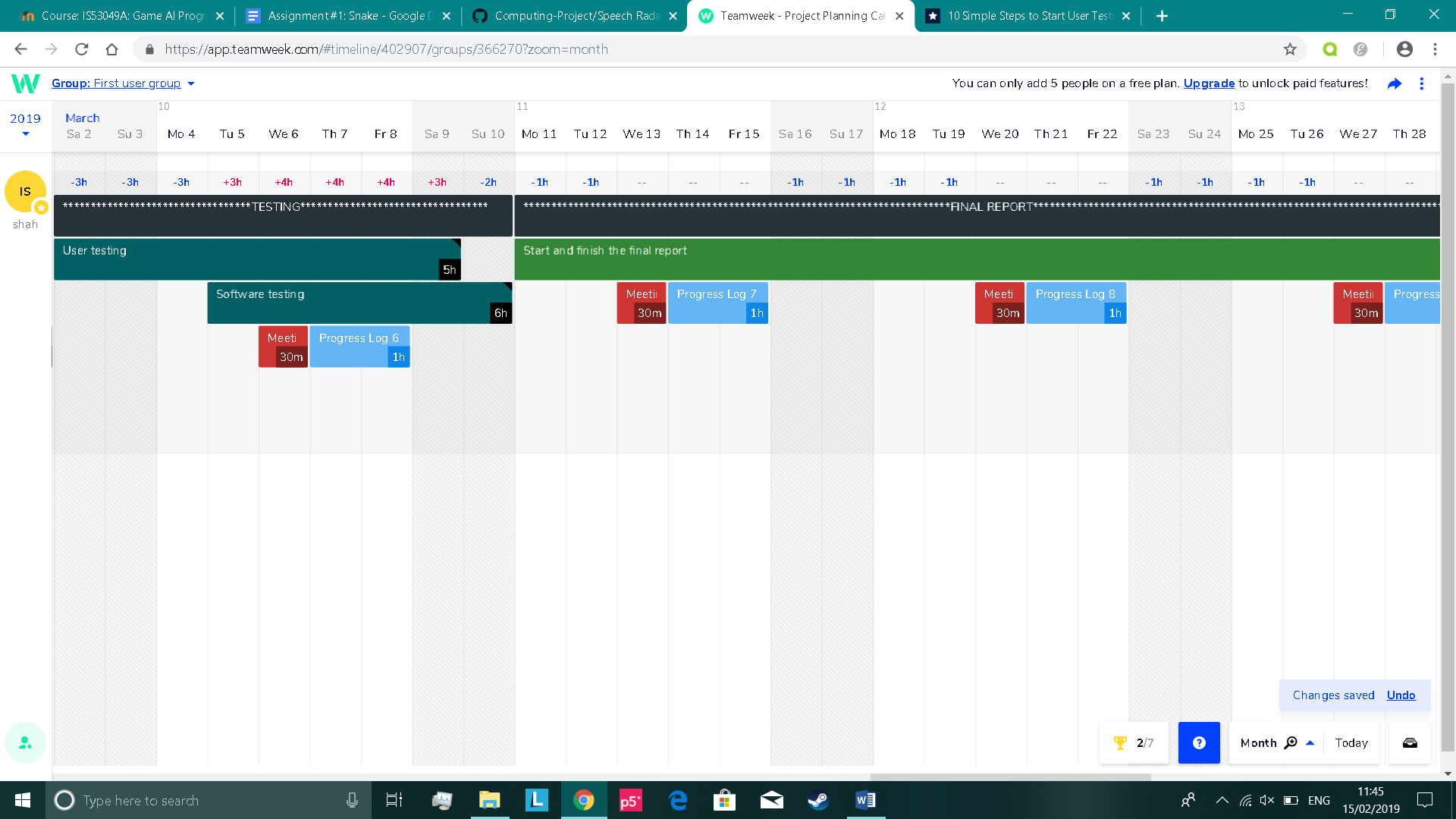
**c. Implementation (continued…)**

**Sub-tasks:**

1. Research on how we can get our app to run in the background and to detect voices
2. Once step 1 is complete and is feasible, implement it
3. If step 2 was successful, then look for a way of preventing the app from draining too much battery power

**Sub-tasks:**

1. Improve the TensorFlow model (e.g. fine-tune hyperparameters etc.)
2. Maybe add more words for recognition
3. Enhance its appearance

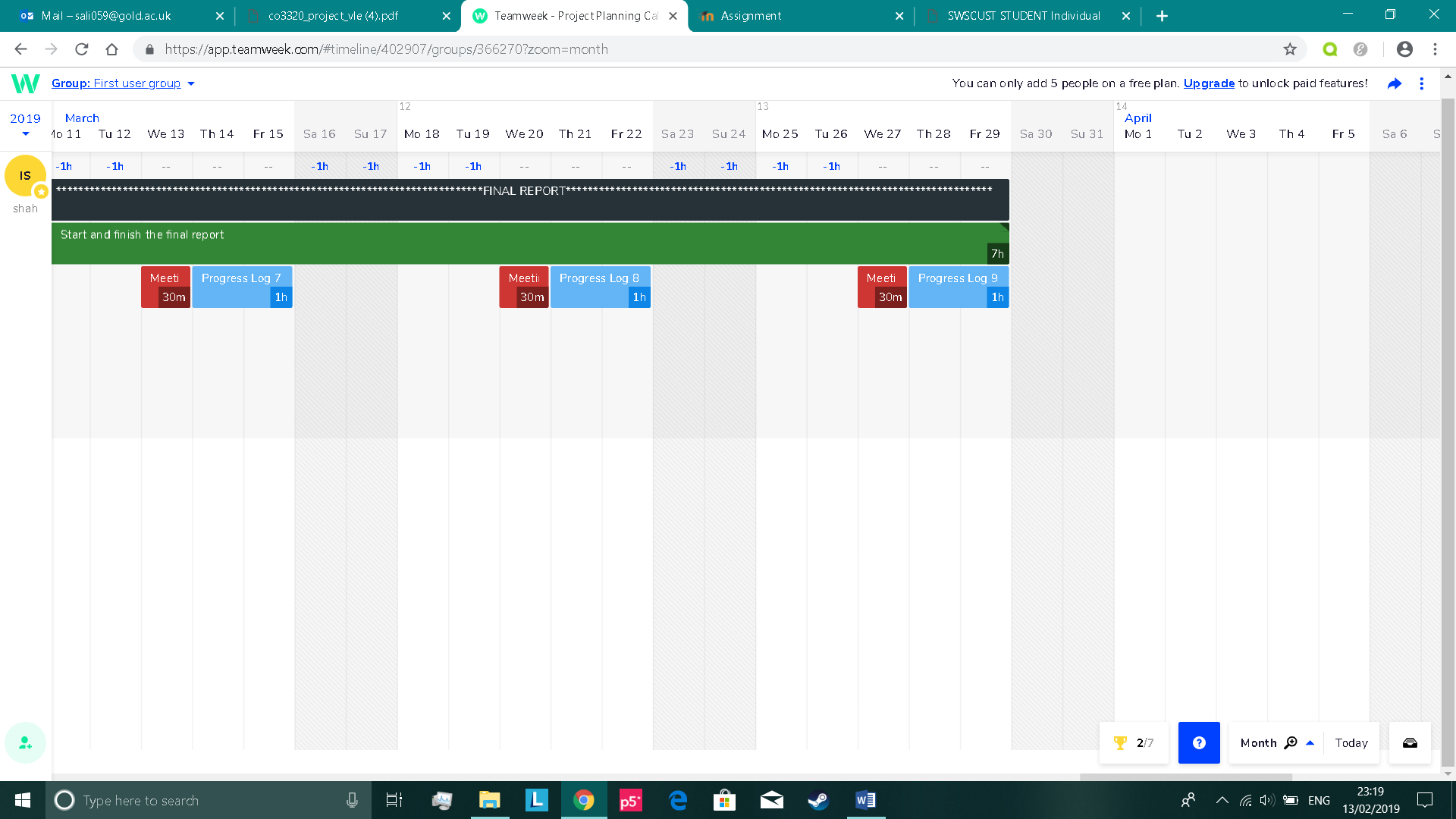
**d. Testing**

Sub-tasks:

1. Cover functional testing
2. Cover non-functional testing
3. Cover static testing
4. Cover dynamic testing

Sub-tasks:

1. Create a survey with questions that require ratings of the app
2. Look for random people in the university, outside university or online and have them test the app and fill the form out
3. Get a large quantity of feedback and make sure to adhere to rules in compliance statement
4. Rectify any issues in the app based on the user feedback

**e. Final report**

**Textual information:**

1. Start off in completing the literature review, methods and results section on the first week
2. Then, complete the discussion section on the second week
3. Finish off with a good introduction in the beginning and a conclusion at the end on the second week

**Visual information:**

1. Add all the appendices and references at the end of the report on the third week
2. Add the title page, contents page etc on the third week

**5. Progress to date**

**a. Design phase**

At the beginning of my project, I began the design phase. With little to no knowledge in what my plan was for this project, it may seem to most people that gaining insight/research should have been what I started doing first. The reason why I chose this order is because I wanted to brainstorm my ideas/thoughts of how I think my app would function and look like. It will be interesting to see at the end what I did in the design phase and compare it to my actual implementation. Just below, are the tasks I did in my first week:

* I created a class diagram to describe the app structure. Can be seen in appendix 7a
* I created an activity diagram to describe the app behaviour. Can be seen in appendix 7b
* I designed the appearance of the app using wireframes. Can be seen in appendix 7c

**b. Research phase**

A paper I read online written by Jansson Patrick helped me to understand how I could convert sound into data. Suppose we have a dataset that contains a set of WAVE files, that are all one-second long. Each file must be represented in a vector with a sampling rate of 16,000. We then extract the features from the raw waveform and convert it into time and frequency domain. This technique is known as MFCC. I was further encourage when reading this paper that I could use a library called Librosa to do this easily in Python. The reference can be found in reference list 8.1

Another paper I read online written by Sanjay Krishna Gouda, Salil Kanetkar, Davan Harrison and Manfred K Warmuth gave me insight into the speech commands dataset. It was publicly provided by Kaggle in order for users to compete in their audio recognition competition. The dataset is an appropriate size for my project and consists of 30 words in total. From this paper, I extended my research even further and discovered that this dataset would be great in detecting accents because it contains thousands of one second audio clips for each word said in numerous accents. It got me thinking that users of my app could utter one of these code words to find their phone. The reference can be found

A reddit forums page I read helped me to understand how I could get the app to work in the background and at the same time, recognise voice. One reddit user said it may be worth looking into an API called Sphinx, which is something I am currently doing. The reference can be found

I devised a broad plan of how to implement the speech recognition screen on Android Studio. The Github link mentioned in references assisted me in creating this plan

**c. Implementation phase**

During this phase, I was able to complete my compliance statement and finish my implementation of the create account screen, where users are able to register and have their details sent to the Firebase database safely. The account passwords are securely hashed, so attackers cannot steal passwords if they tried to gain unauthorised access to the backend. Users are also expected to verify via email for authentication. I started my implementation at an early stage in the attempt to gain more time to spend for implementing the speech recognition screen. Can be seen in appendix 7d

I finished my implementation of the login screen where the input fields check whether the account exists in Firebase, the information inputted is incorrect and if the email address has been verified by the user. Can be seen in appendix 7e

I began my implementation of the speech recognition screen not too long ago. I have had a few setbacks here and their, but I am confident I can finish on time and ensure it functions the way I want it to. So far, I have designed the screen and got it to recognise three words (‘cat’, ‘bed’ and ‘happy’.) Can be seen in appendix 7f

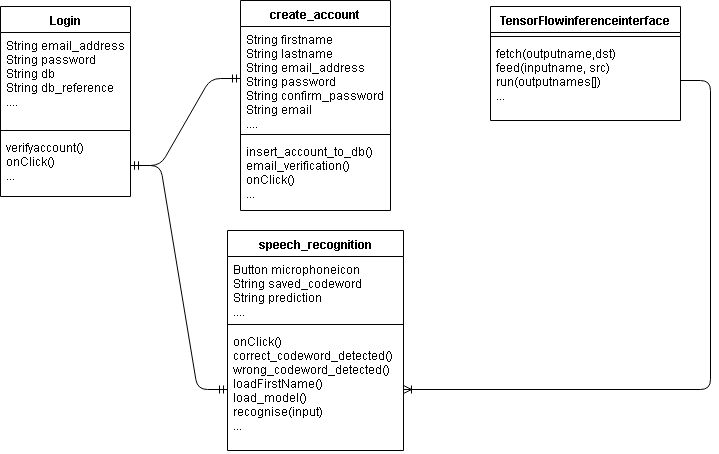
**d. Summary**

I think I completed all tasks I set out to do up until this point with a few minor setbacks here and their in which I’ve overcome. One of the key tasks that I’m currently pleased in doing is gaining insight from research of how to work with sound and converting it to recognisable data. This helped me to gain knowledge and understand the logic behind how the input is generated for my TensorFlow model.

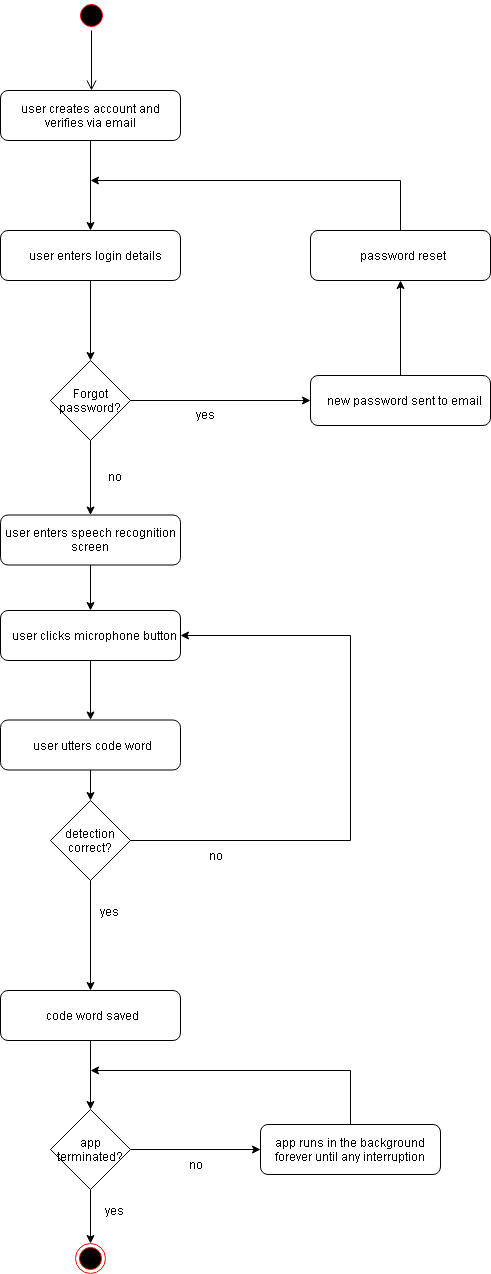
**6. Planned work**

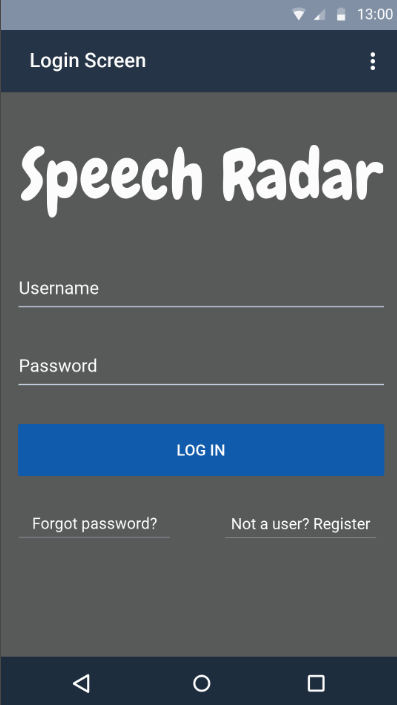
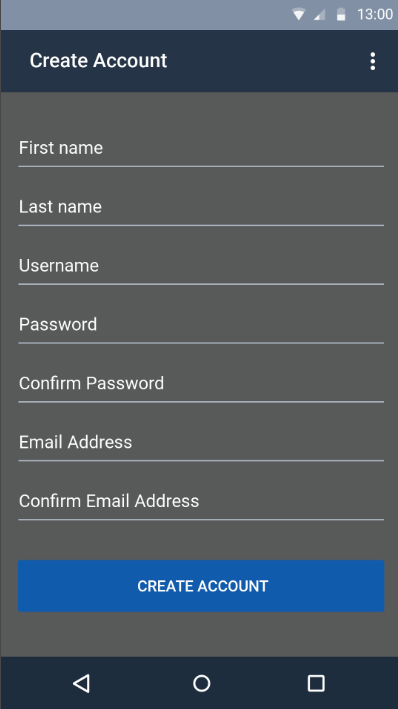
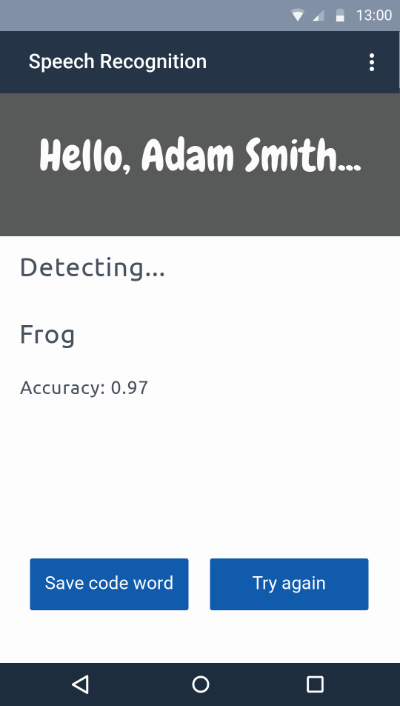
Based on what I mentioned in the previous section, I completed the design and research phase of my project. Now I am in the implementation phase and have created a quarter of my final application on Android Studio. I have increased the amount of time and work I do drastically on a daily basis. What I need to do next is to finish off the implementation to quickly get into the testing phase where we will begin user testing and software testing (e.g. functional, non-functional etc) for quite some time. If this all goes to plan, then I should have enough time for the final report at the end.

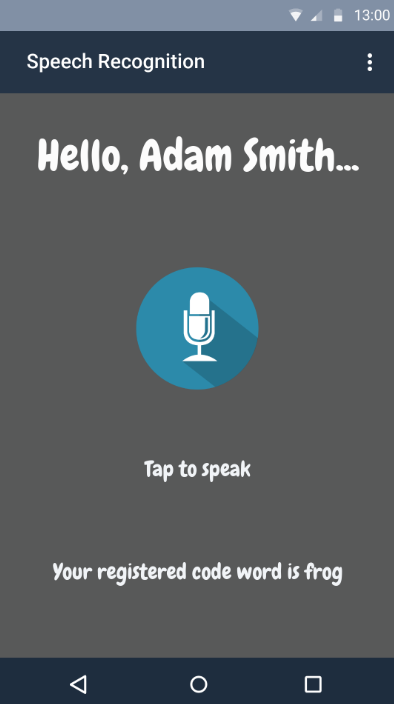
**7. Appendices**

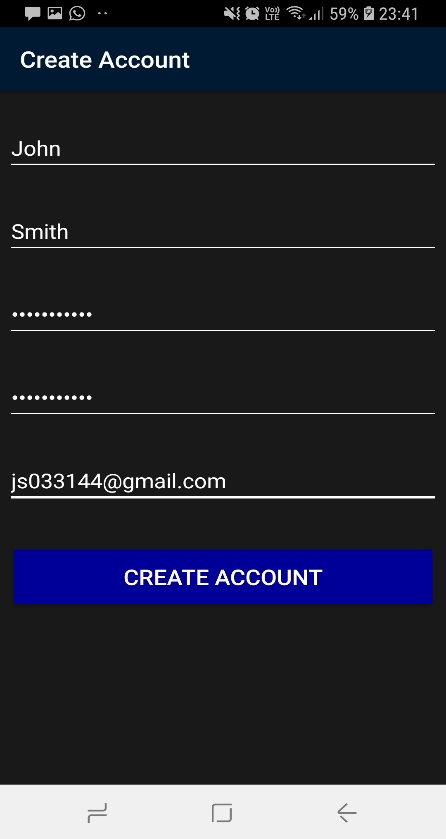
**a. Class diagram**

**b. Activity diagram**

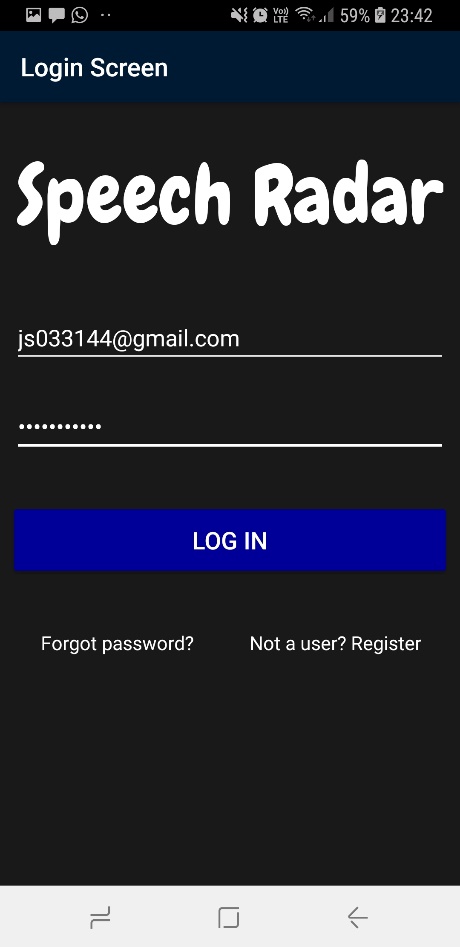


**c. Wireframes**

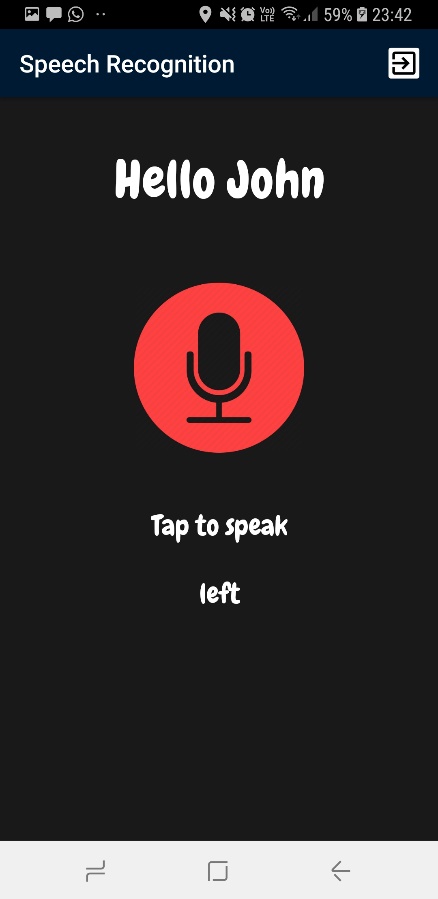


**d. Create account screen**

**e. login screen**



**f. speech recognition screen**



<https://www.reddit.com/r/androiddev/comments/r7zea/voice_recognition_background_service_possible/>

<https://www.theseus.fi/bitstream/handle/10024/144982/Jansson_Patrick.pdf?sequence=1&isAllowed=y>

<https://arxiv.org/pdf/1803.03759.pdf>

**EXTRA NOTES**

* Make sure to mention that it makes sense to work with a small set of words, as our aim is to have users register code words, not just any other words like for example “the.”