SCHOOL OF ELECTRICAL ENGINEERING AND TELECOMMUNICATIONS

Thesis B Progress Report

By

Madeline Mariam Younes

Student ID: z5208494

Thesis Submitted as a requirement for the degree Bachelor of Engineering (Electrical Engineering)

Submitted: June 30th, 2022

Supervisor: Dr Beena Ahmed

Table of Contents

1. Progress Update

1.1 Summary of Thesis A

The aim of this thesis is to investigate the transfer learning as a novel and effective method of creating an accurate Arabic Dialect Identifier. Thesis A was focused on building basic background knowledge through literature reviews, meetings with my thesis supervisor Dr Beena Ahmed and with a contact in industry Ian Thorvaldson from Dubber Ai. During thesis A, existing methods for designing language identifiers and Arabic dialect identifiers were researched. As well as transfer learning methods, possible pretrained models to use, machine learning models used for classification tasks and Arabic dialect datasets. The preliminary conducted during Thesis A mainly consisted of basic data analysis and setting up the development environment to be used in Thesis B & C.

1.2 Thesis B Progress

The following is a breakdown of the progress made since Thesis A, during term 2 of 2022.

1.2.1 Week 1

The focus of week 1 was on creating a python script to generate a text file with a list of the audio files and their dialect labels to be used for training. The ADI17 dataset is labelled using 17 regional dialectal labels, the initial focus of the thesis is to develop a system that is able to work on 4 broader umbrella dialectal groups comprised of those 17 groups. The python script takes in whether the labels should be from the regional or umbrella dialect and the amount of time to be collected from each dialect. Then ensures there is a random set of audio files fulfilling those requirements listed in the text file. Sometime during Week 1 was also spent on installing required python libraries and reading hugging face tutorials on data preparation and using a pretrained model for audio classification.

1.2.2 Week 2

There were some errors in the text file generating script, that were debugged during this week. Also a basic filtering and normalisation script was written to pre-process the audio. The pre-processing methods and their result are further discussed in *Section 1.4*. Also, during week 2 there was an attempt to use the hugging face tutorials to implement a basic model. Access to katana was requested and granted.

1.2.3 Week 3

During this week an alternate filtering method using a Butterworth bandpass filter was experimented with as a way to pre-process the data. I read through some of Reenè's documentation on the recommended working environment for using katana and training using wav2vec. I also attended a workshop run by Reenè on working with katana and creating a wav2vec implementation. At this workshop I was able to ask questions and gain a better understanding of different data preparing methods that can be such as batching to improve training efficiency and was able to gain a better understanding of how to use wav2vec. I also researched a potential benchmark model to use but was unable to find any code for language identification or dialect identification that used pretrained models.

1.2.4 Week 4

Based on workshop adapted an example script to be used for an audio classification task like dialect identification. I also noticed that the input file should be a csv instead of text file, so my file generation script was changed to output a csv rather than a text file. It was also recommended to download the dataset onto katana. The ADI17 dataset is quite large with the training data portion being 300GB in compressed form. The entirety of the dataset stored on a Linux computer I can access through ssh-ing. On Kanata I was allocated around 15GB of storage. The dev and test portions of the dataset were around 2GB each and so the space provided was sufficient for those files to be stored on katana. To use the training portion of the dataset, I think a method of batching will have to be used where portions of the dataset are downloaded onto katana, then deleted to make room for new sets of data. Some attempt of a script to port between katana and the Linux computer but was sidelined before being at the point of fully functional. The dev portion of the dataset will be used for initial training and experimenting with the wav2vec model.

1.2.5 Week 5

Attempted to train the model designed in week 4 ran into some errors in the code particularly in the data processing stages where it converts the audio files into arrays. The majority of this week was dedicated to fixing errors related to running the training model.

1.3 Pre-processing Filter and Normalisation

The ADI17 dataset is sourced from scrapped YouTube clips part of the preprocessing explored were different methods to removing the noise component. Two audio files from the dataset were used for testing, Audio 1 contained background noise in the forms of bird noises and a baby crying, while Audio 2 contained instrumental background music in addition to the speaker. The first method tested was using the noisereduce python library. The library's algorithm utilises both stationary and non-stationary noise reduction. Stationary noise reduction maintains a constant estimate for the noise threshold across the whole audio file whilst the non-stationary updates its estimate overtime. The library effectively reduced constant noise such as in *Audio 2* as seen in *Figure 2* without distorting the voiced audio. Although, it struggled with *Audio 1* shown in *Figure 1*, it filtered a lot of the low volume environmental noise but struggled with the louder noise of the baby crying. When passed through the noisereduce function, the voiced audio became distorted. A butterworth bandpass filter with the cutoff bands at the frequency ranges of voiced audio between 50Hz and 500Hz was also tested. Although, it didn't distort the voiced audio but was not observed to be highly effective in removing noise from either signal. More experimentation will be conducted and the use of pre-processing will be further assessed during this thesis.

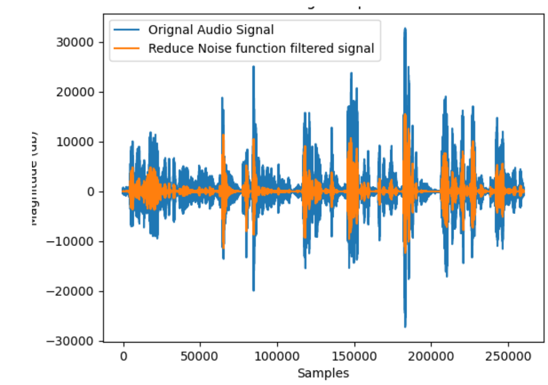

Figure 1: Audio 1 filtered using reduce noise library

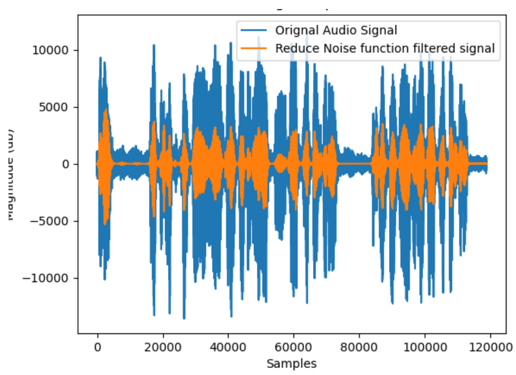

Figure 2: Audio 2 filtered using reduce noise library

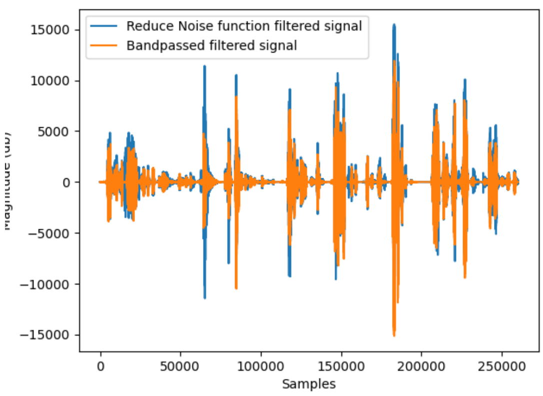

Figure 3: Audio 1 filtered using Butterworth bandpass filter

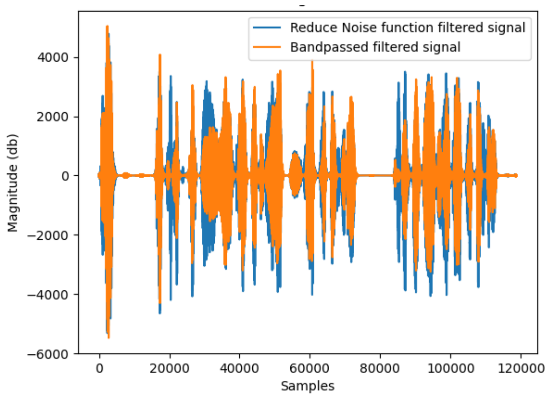

Figure 3: Audio 2 filtered using Butterworth bandpass filter

2. Reflection

Over thesis A and half of thesis B, I have gained a greater understanding of transfer learning, how to utilise pretrained models for specific task, about language identification systems and dialectal identification systems. I have gained an understanding of the challenges that I will be facing throughout the rest of my thesis, particularly the constraint on time as well as how long it will take me to make changes and bug fixes. As well as this I have gained a vital skillset in self directed project management, communication and the process of researching a topic.

My initial concern when undertaking thesis was finding a topic I could find interesting enough to explore over a year. Through thesis A I was able to establish an area to research that I found interesting, addressed a real need and was excited to work on. The greatest challenge I have had so far in thesis B has been translating idea to implementation. I found the initial process of translating research I had conducted in Thesis A to code a daunting and overwhelming. Particularly, in terms of knowing how to start structuring my systems to run on katana and my training scripts. Having the workshop with Reenè, really helped me get past this initial fear and from the knowledge I developed over the session I have almost been able to finetune an initial basic wav2vec model.

Through this thesis I have developed my ability to work on a large open ended task that is primarily self directed. My past project work experiences such as on my summer internship at Dolby or working on the Electrical team of the BluSat rover or the UNSW Illuminae vivid installation were on a shorter timeline of around 8-10 weeks. Also, on all those teams many of the design decisions were primarily made by my superiors or informed based on the knowledge of other people on the team. I have found the longer timeline compared to shorter projects a little intimidating at first, often struggling to decide which portion of the project I should focus. I have found it much more vital to breaking down large tasks into specific smaller subtasks vital to maintaining focus and ensuring that tasks are being completed in a timely manner and. During this thesis, I have primarily been making decisions on the tasks that need to be completed, my approach to the problem and what aspects to focus on. I have found this responsibility and freedom both a challenging and empowering experience. Through learning how to plan, manage my own tasks and solve the problems that arise I have become more resilient and confidence in my competence as an engineer.

Even though I would have liked to have completed more of the implementation by this portion of my thesis I have developed my the baseline technical skills needed to complete this thesis. As well having completed the setup and a lot of the groundwork which will streamline the design process throughout the rest of my thesis. As I become more comfortable with the libraries and development setup I am using I hope that I will be able to work through tasks more quickly. So, I will be able to be more efficient in the later portion of Thesis B and in Thesis C.

Overall, through my thesis so far I have development both my technical and 'soft' project management skills.

3. Revised Plan

Since, starting Thesis B the initial step of getting an initial model trained has taken longer than expected. More time than expected was needed for familiarising myself with the huggingface libraries and preparing the data. The updated gantt chart factors in this extended time taken on the initial setup. Also, wav2vec implementation was prioritised for the initial experimentation and so this is reflected in the revised gantt chart.

3.1 Updated Gantt Chart

Tasks	Weeks	Completion (%)
Preliminarly Reaserch: Self supervising models, ASR tech, LIDs, DIDs, Current methods, Arabic Dialects etc.	4	100
Find a relevant dataset	1	100
Perform basic data analysis on the chosen dataset	2	100
Set up the SSH computing and become familiar with utlising the Gadi computing cluster at NCI.	2	100
Prepare for Thesis A Seminar	2	100
Familarise using wav2vec	2	10
Thesis A Final Report	2	100
Dataset Preperation	1	100
Implement XLS-R with basic pooling + linear layer	2	70
Test accuracy of various audio lengths with 100hrs of training on downstream model	2	0
Implement CNN/BiLTSM Linear Downstream Model	2	0
Implement with wav2vec 2.0 pretrained model	2	90
Test accuracy using various amounts of training data	2	0
Thesis B Report	3	0
Review of Expected Outcomes	2	0
Makes any changes that need to be made to the model or the focus of the thesis	1	0
Train using fine grain dialect classification groups	2	0
Test accuracy of both options of pretrained and downstream models	2	0
Test using codeswitch input data	2	0
Final verification and review	2	0
Final Report	3	0
Final Poster	2	0

3.2 Risk Management

The major risk for this project is having insufficient time to complete the listed tasks due to road blocks in the development process or due to the time needed for training. To mitigate this risk, I will need to effectively manage my time and effectively prioritise tasks that need to be completed for the thesis. I will have regular check-ins with my thesis supervisor Dr Beena Ahmed, Ian from Dubber and Jack Murray another student working on a similar thesis to ensure that I am progressing well through my thesis.

In terms of technical risks fine tuning the model's hyper parameters efficiently maybe a time intensive process. To increase the efficiency of this process I plan on writing a script to cycle through testing and assessing the best value to use for each metaparameter. Reené has also provided a document describing each metaparameter and it's use, while her thesis was on a ASR use case this understanding can also be applied to an audio classification task like Arabic Dialect Identification.

Another risk is managing the training portion of the ADI17 dataset, as it is so large in storage size. Batching may provide a good method of still being able to use this large set of data in smaller more manageable chunks. Although, it still needs to be assessed if a model can be stored midtraining to allow for a chunk of the data to be downloaded. It also may be more efficient to use the training dataset as unlabelled for self supervised learning and the dev portion for supervised learning.

There are a lot of unknown aspects to the ‘nitty gritty’ implementation of this thesis and there is a risk I could get stuck on a bug or broken code. To reduce the time spent debugging or becoming frustrated when encountering roadblocks, I will ask for support when I need it. This support and feedback would come from my fellow thesis students and my thesis supervisor.