Engineering Notebook

Tahmina Tisha
Cs 490
Dr. Sultana & Dr Akbas

Tisha 9/21 Notes

Multiple RVS for a certain event

- Conditional probability (given the previous text, then the probability of the next)
- Bayesian theory (p(a|b) to (p(b|a)
- Markov chain (p(s2|s1, s0) = p(s2|s1)

Hidden Markov model (HMM)

- The system described by a state
- We can't observe the state
- The state transition is backward Manner
- Each state can go forward
- Each state transmits observable RV

Use observable RV to infer the state of the HMM

- Phones and tri-phones
- Phones are used to model the pronunciation of words in an utterance
- Speech recognition is can be reformulated as phone recognition
- We need to train the system to recognize different phones
- These phones come from the lexicon dictionary
- To better model the transition of the phones, we use tri-phones

Using HMM to match the time-varying signals

Four states

- start and an end
- Three states for normal sounds
- Five states (forgot what he said)
- Only goes forward
- Cannot be observable

- Count for different accent
- The model probability for the distribution

Two probability models to train

- each hmm has several states to better model changes in pronunciation
- We need to train the probability model that describes the transition from one state to another using the training dataset
- For a given state, the probability distribution of MFCC is described by using GMM
- We need to train this GMM using the training dataset as well. For simplicity, we assume the element
- We assume the elements of the feature vector are independent
- This is supported by using MFCC
- FIRST PHASE: Monophone
- Second phase: tri phones

Decision Tree and Senone

The total number of tri phones is too big
We do not have the resources to use all the tri phones
Decision trees are used to significantly shrink useful tri-phones
Decision phones can be formed by linguists. They can also be learned, which is especially important to the special applications or language of small speaker
Each kept tri phones is expressed using HMM

To Do Kaldi Product Backlog Sprint 1

SRS V1, Due Sep 28, 10 hours, Tabitha Hudson, Milan Haruyama, David Serfaty, Maxwell Moolchan, Tahmina Tisha

- Introduction
    - Purpose
    - Document Conventions
    - Intended Audience
    - Product Scope
    - References
- Overall Description
    - Product Perspective
    - Product Functions
    - User Classes
    - Operating Environment
    - Design and Implementation Constraints
    - User Documentation

        - Assumptions and Dependencies
    - External Interface Requirements
        - User Interfaces
        - Hardware Interfaces
        - Software Interfaces
        - Communications Interfaces
    - System Features
        - Systems Features
    - Other Nonfunctional Requirements
        - Performance Requirement
        - Safety Requirement
        - Security Requirement
        - Software Quality Attributes
        - Business Rules
    - Other Requirements
        - Other Requirements
        - Appendix A
        - Appendix B
        - Appendix C

SDS V1, Due Sept 28, 10 hours, Tabitha Hudson, Milan Haruyama, David Serfaty, Maxwell Moolchan, Tahmina Tisha

- Introduction
    - Purpose and Scope
    - Project Executive Summary
    - System Overview
    - Design Constraints
    - Future Contingencies
    - Document Organization
    - Project References
    - Glossary
- System Architecture
    - System Hardware Architecture
    - System Software Architecture
    - Internal Communications Architecture
- Human-Machine Interface
    - Inputs
    - Outputs
- Detailed Design
    - Hardware Detailed Design
    - Software Detailed Design
    - Internal Communication Detailed Design
- External Interfaces
    - Interface Architecture

- Interface Detailed Design

SDD Revision History

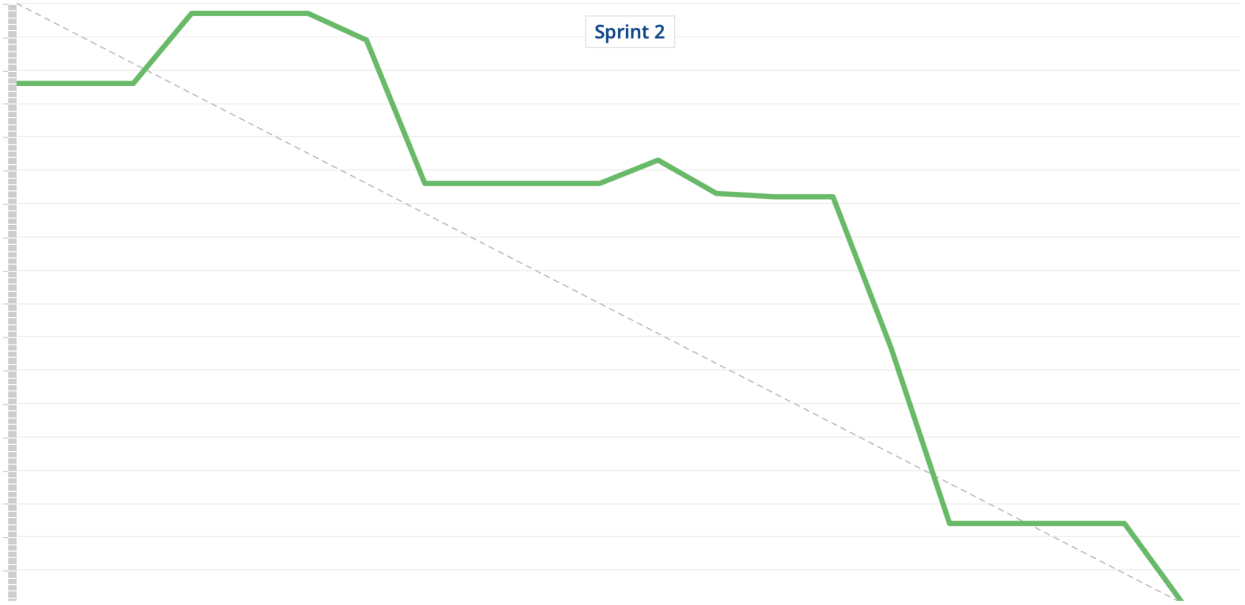
Revision History

Name	Date	Reasons For Change	Version
Tabitha, Milan, David, Max, Tisha, Adam	09/29/2023		V1.0
Tabitha	10/24/2023	Writing Section:1.2.1	V2.1
Tisha, Tabitha, Milan	10/24/2023	Rewriting the section: 2.2	V2.2
Tabitha, Tisha, Milan	10/24/2023	Writing the section, Rewriting, and editing: 1.2	V2.3

SRS Revision History

Revision History

Name	Date	Reason For Changes	Version
Tabitha, Milan, Tisha, David, Adam, Max	09/29/23	Starting the document	V1.0



Notes

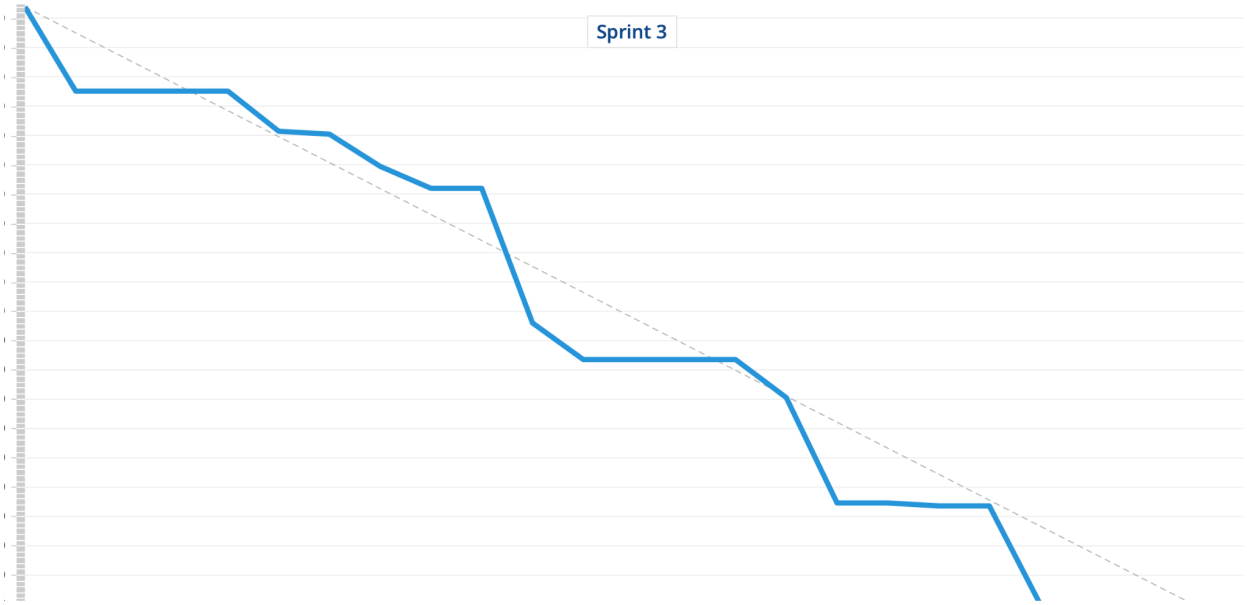
We all started working on the requirements. We took notes in class and implemented and corrected that for SRS final version.

SDD Revision

Tisha	10/28/2023	Asking TA: 2.1 Writing Section: 2.1, 5.1	V2.9
Tabitha	10/29/2023	Writing/Rewriting Sections: : 1.2.1, 2.1, 2.2, 3.1, 3.2, 4, 4.1, 4.2, 5.2	V2.10
David	10/29/2023	Writing/Rewriting Sections: 1.2.1, 1.2.2, 1.2.3, 2.1. 2.2, 3.1, 3.2, 4, 4.1, 5, 5.1, 6	V2.11
Tisha	10/29/2023	Writing/Rewriting Section: 2.1	V2.12

SRS Revision

Tisha	11/05/2023	Editing: 2.4, 2.5	V3.5



Tisha	11/05/2023	Edited Section 2.2	V3.3
Adam	11/07/2023	Writing/Rewriting: 3.1, 3.2	V3.4
Tisha	11/07/2023	Writing/Rewriting 3.1	V3.5
Adam	11/07/2023	Writing/Rewriting 3.1	V3.5
Milan	11/07/2023	Editing all Sections, reformatting Table of Contents	V3.6
David	11/11/2023	Editing and Rewriting: 4.2 Updating DFD model	V3.7
Tabitha	11/11/2023	Editing and Rewriting: 4.2	V3.8
Tisha	11/11/2023	Editing 4.1	V3.9

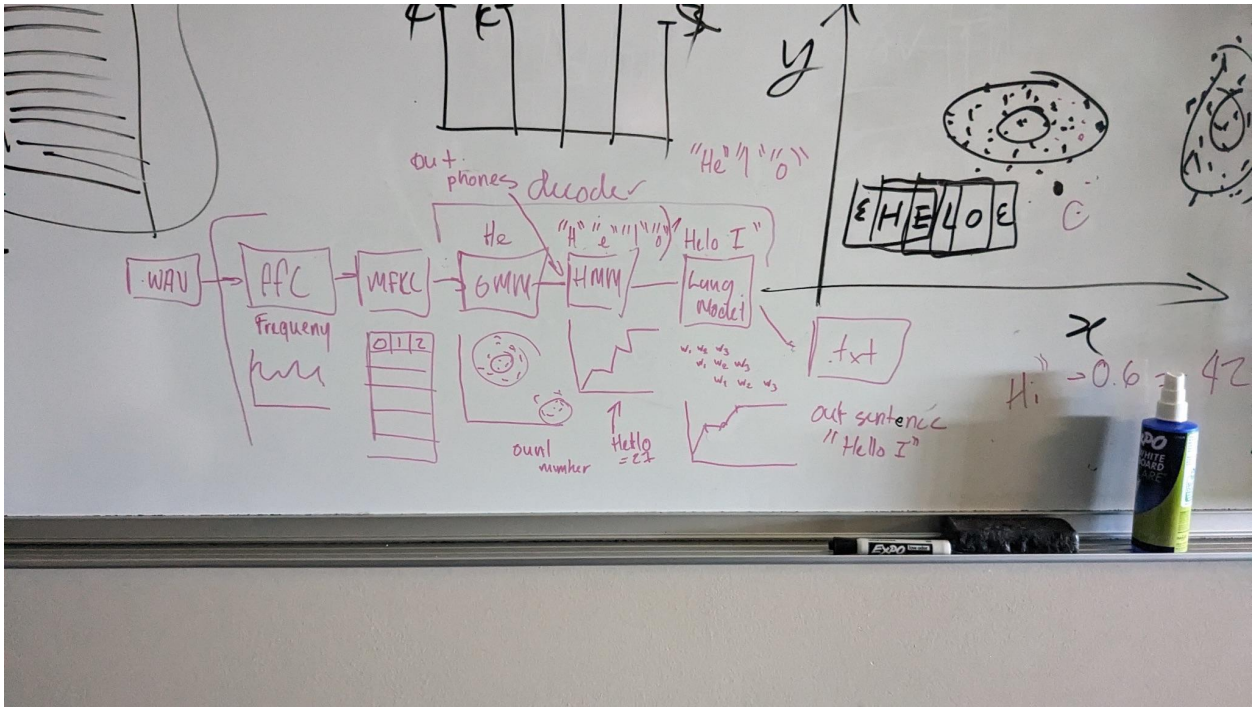
Tisha	11/16/2023	section 4.1 (needs to be checked)	V3.14
Tabitha	11/16/2023	Reading and Commenting Sections : 2-5 For accuracy to the current requirements Update Classes Diagram Update/Rewrite Section: 2.1	V3.15
David	11/18/2023	Editing and rewriting Use Cases Remaking Use Case Diagram V2.1.1 Making Use Case Diagram V2.2.2 Editing DFD V3.1.3, V3.2.2, V3.3.1 Rewriting and editing 4.1	V3.16
Tabitha	11/18/2023	Editing and rewriting Use Cases Remaking Use Case Diagram V2.1.1 Making Use Case Diagram V2.2.2 Editing DFD V3.1.3, V3.2.2, V3.3.1 Rewriting and editing 4.1	V3.17

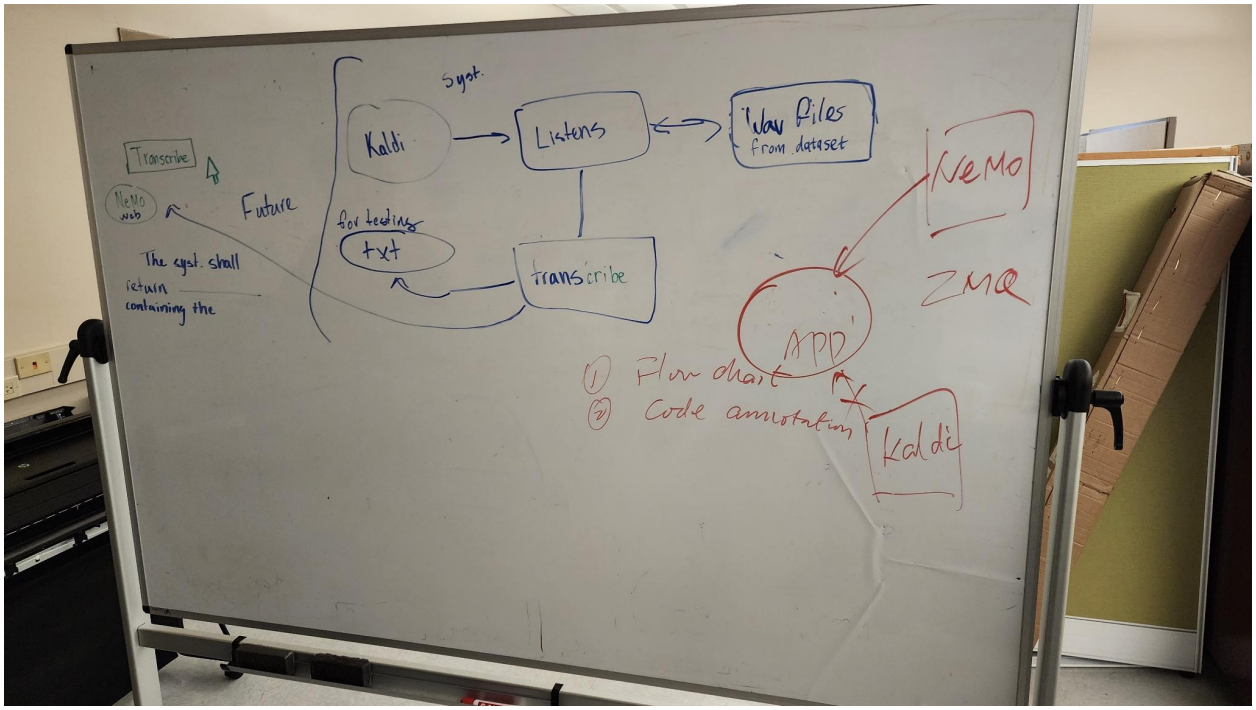
Tisha	11/18/2023	Rewriting section 5.2 (needs to checked for accuracy)	V3.18

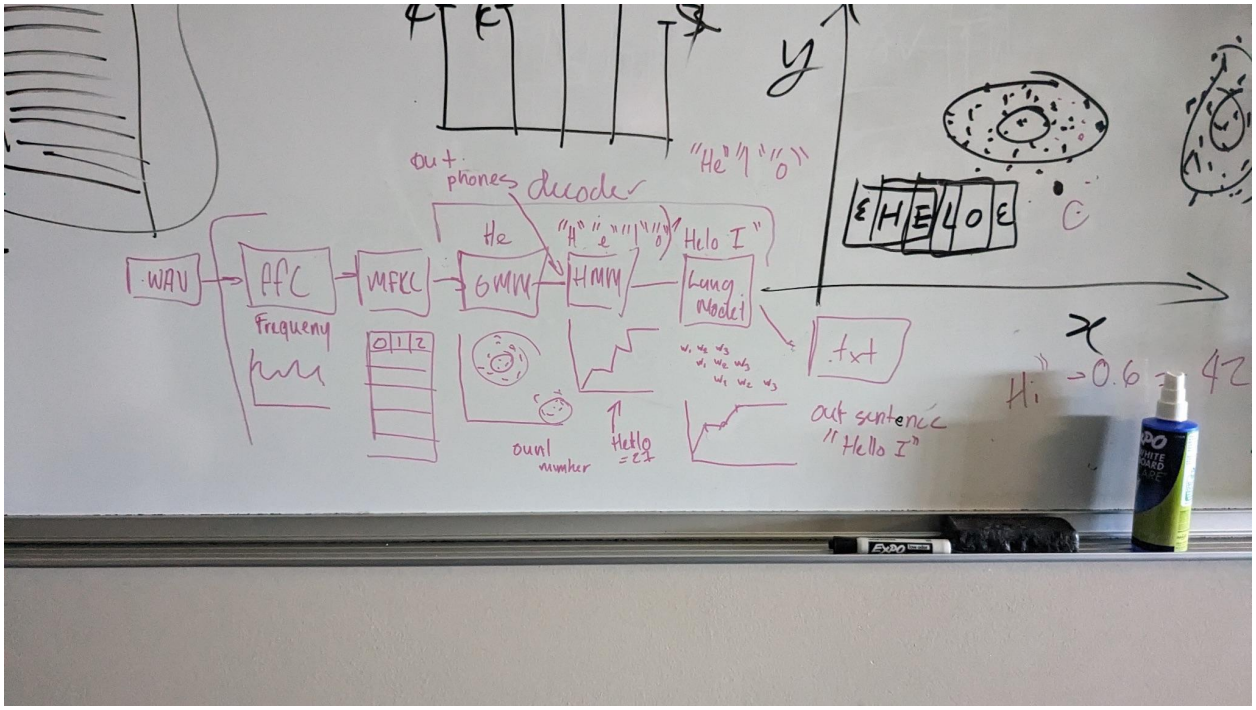
SRS Revision

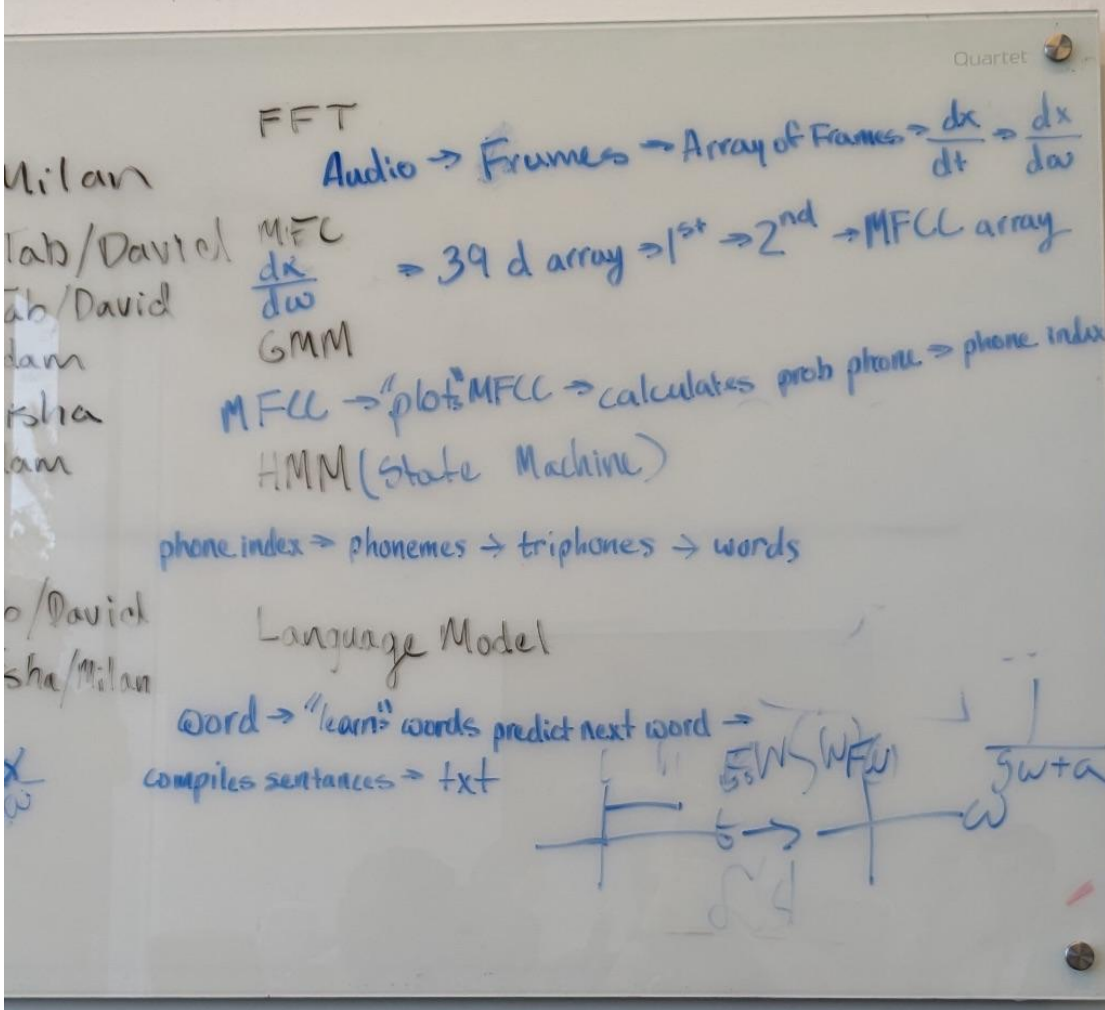
Tisha	11/15/2023	Editing section: 5	V3.15
Tabitha	11/15/2023	Editing: 1.5, 2.3, 4.1, 5.1	V3.16
Tabitha	11/16/2023	Review/Edition/Commenting Section: 4 Update Class Diagram Update/Review Section: 2.3	V3.17
Tisha	11/18/2023	Rewriting section 5.2	V3.18
Tisha	11/18/2023	Edited section 5.2	V3.19

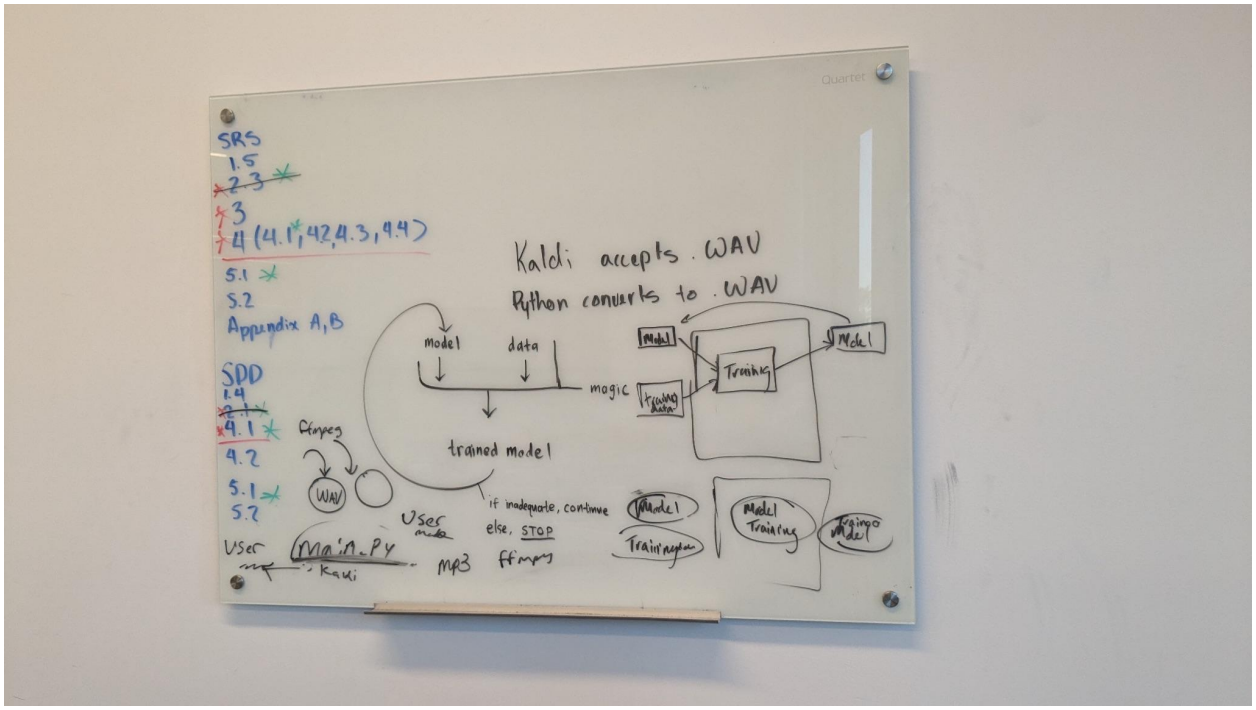
Photos Throughout the Semester

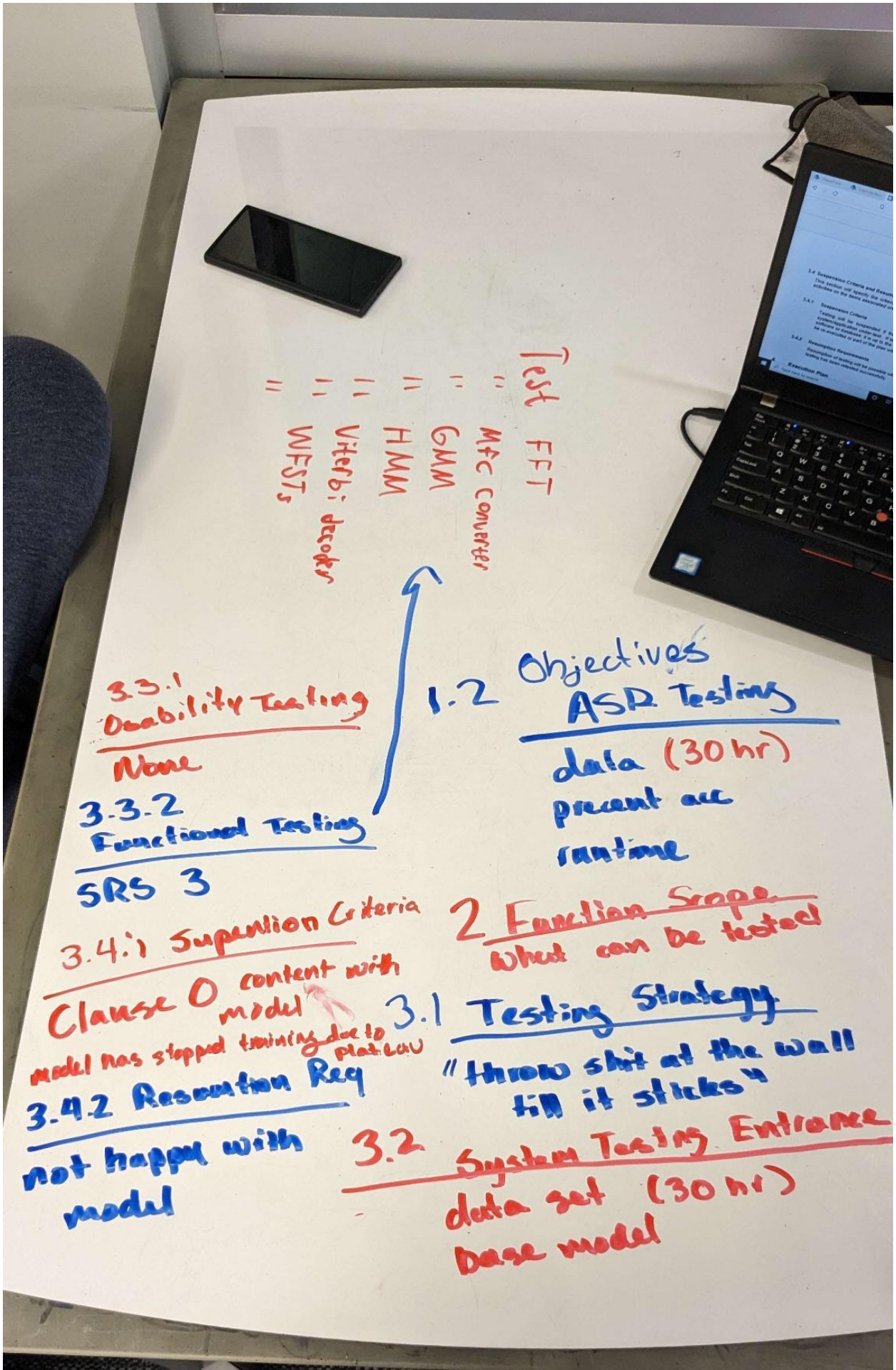


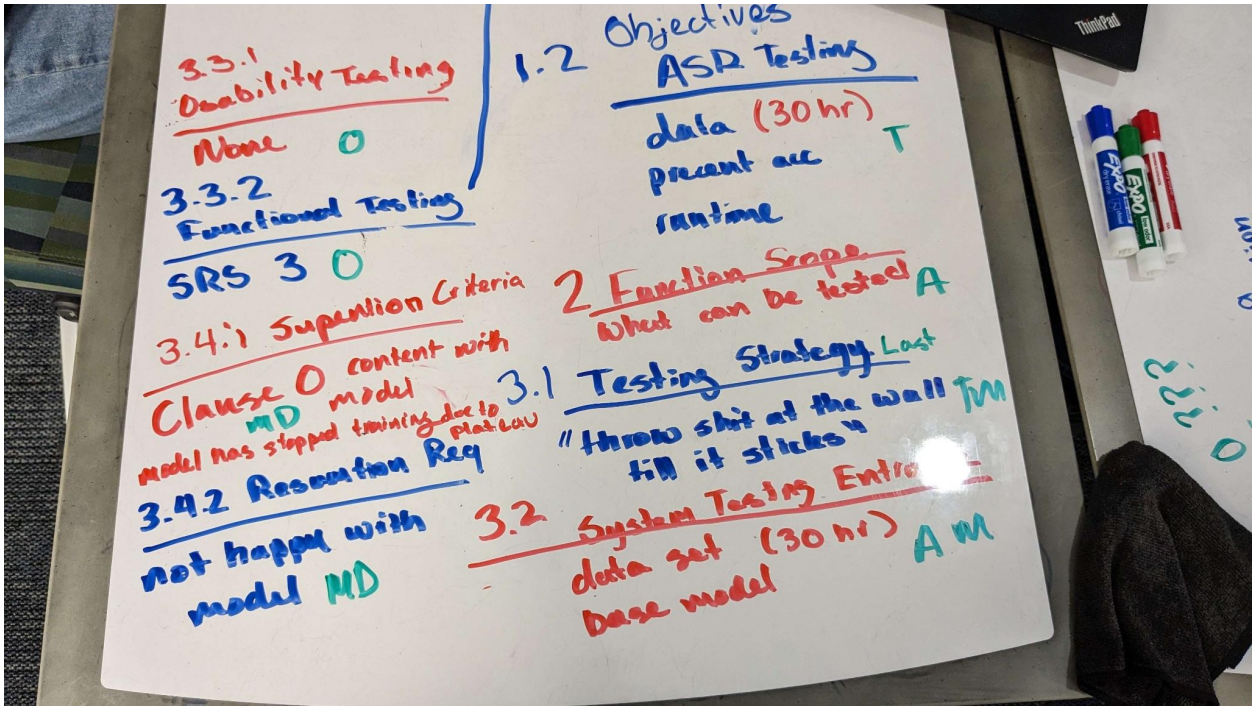

Syst.
Kaldi
Listens
.wav files
from dataset
Transcribe
NeMO web
Future
for testing
.txt
transcribe
The syst. shall
return
containing the
NeMO
ZMQ
APP
1 Flow chart
2 Code annotation
Kaldi


WAV
FFC
Frequency
MFCC
He
GMM
HMM
"H" "e" "l" "l" "o"
Helo I"
Lang Model
.txt
out phones
decoder
"He" "l" "o"
out sentence
"Hello I"
Hello
=27
count
number
y
x
Hi
-0.6
42


Quartet
FFT
Audio → Frames → Array of Frames → dx/dt → dx/dw
Milan
lab/David
ab/David
dam
isha
am
MFC
dx/dw
→ 39 d array → 1st → 2nd → MFCC array
GMM
MFCC → "plot" MFCC → calculates prob phone → phone index
HMM (State Machine)
phone index → phonemes → triphones → words
o/David
sha/Milan
Language Model
word → "learn" words predict next word →
compiles sentances → txt
y=w+a


SRS
1.5
2.3
3
4 (4.1, 4.2, 4.3, 4.4)
5.1
5.2
Appendix A, B
SDD
1.4
2.1
4.1
4.2
5.1
5.2
Kaldi accepts .WAV
Python converts to .WAV
model
data
magic
trained model
Training
Model
ffmpeg
WAV
User
mp3
if inadequate, continue
else, STOP
Model
Training
Model
Training
User
Main.py
Kaldi


Test FFT
= Mfc converter
= GMM
= HMM
= Viterbi decoder
= WFSTs
3.3.1 Usability Testing
None
3.3.2 Functional Testing
SRS 3
1.2 Objectives
ASR Testing
data (30 hr)
precent acc
runtime
3.4.1 Supention Criteria
Clause 0 content with model
model has stopped training due to plateau
2 Function Scope
What can be tested
3.1 Testing Strategy
"throw shit at the wall till it sticks"
3.4.2 Resumtion Req
not happy with model
3.2 System Testing Entrance
data set (30 hr)
base model


3.3.1
Usability Testing
None O
3.3.2
Functional Testing
SRS 3 O
3.4.1 Suspension Criteria
Clause O content with
MD model
model has stopped training due to plateau
3.4.2 Resumption Req
not happy with
model MD
1.2 Objectives
ASR Testing
data (30 hr)
present acc
runtime
T
2 Function Scope
What can be tested
A
3.1 Testing Strategy Last
"throw shit at the wall
till it sticks"
TM
3.2 System Testing Entr
data set (30 hr)
base model
A M
ThinkPad
EXPO
EXPO