Automated Speech to Sign Language Conversion using Google API and NLP

CS-19

Ritika Bharti-20154062

Sarthak Yadav-20154004

Vishnu Kumar-20154122

Sourav Gupta-20154075

Nenavath Sirisha-20154083

UNDER THE GUIDANCE OF

Dr. RAJITHA BAKTHULA

Motivation

- 1.3 million people with hearing impairment in India.

  - *2011 Indian Census*

- Roughly 1 percent of Indian Population i.e. 18 million are deaf.

  - *National Association for Deaf*

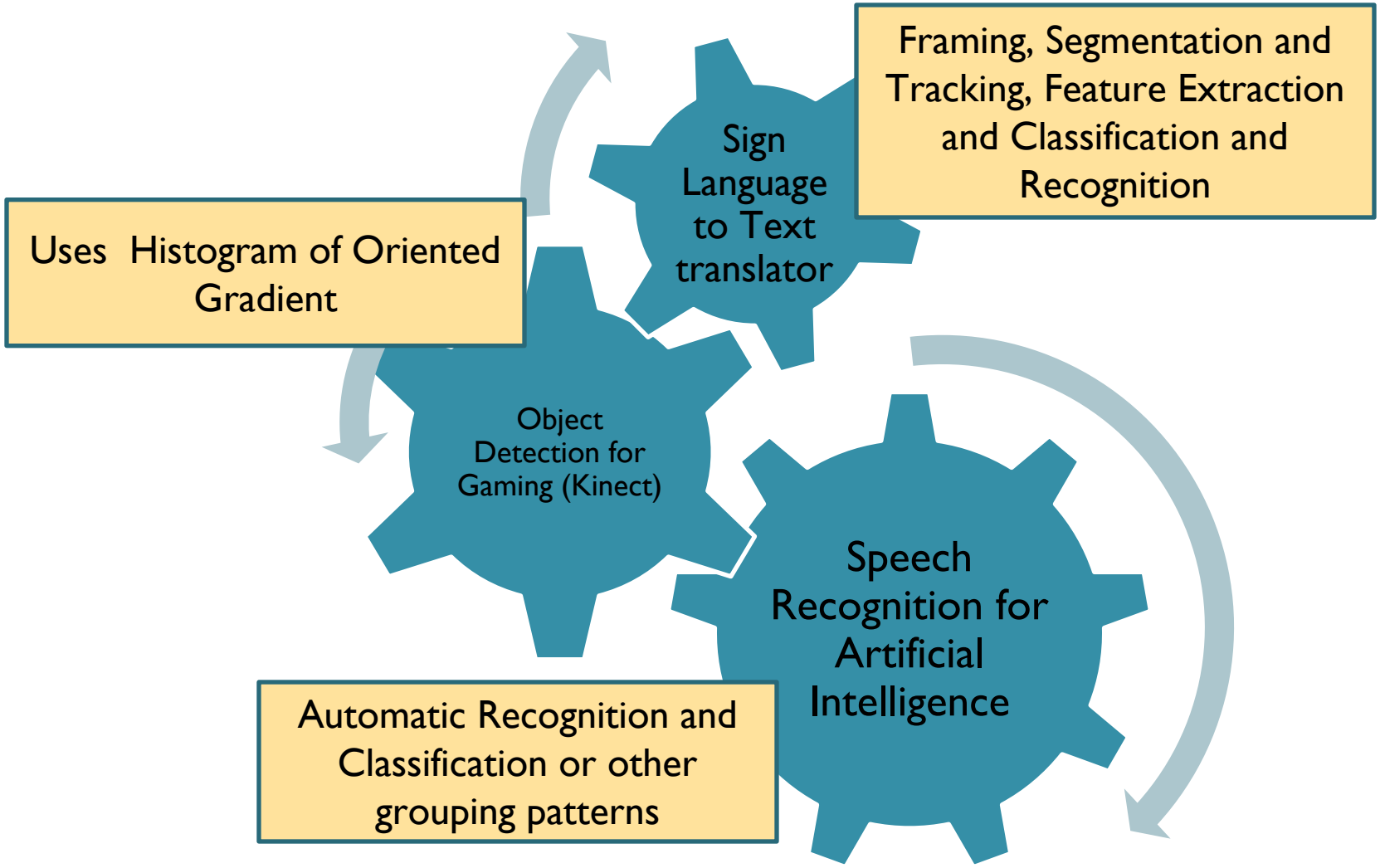
- 250 certified sign language interpreters translating for a deaf population of between 1.8 million and 7 million.

  - *GlobalPost*

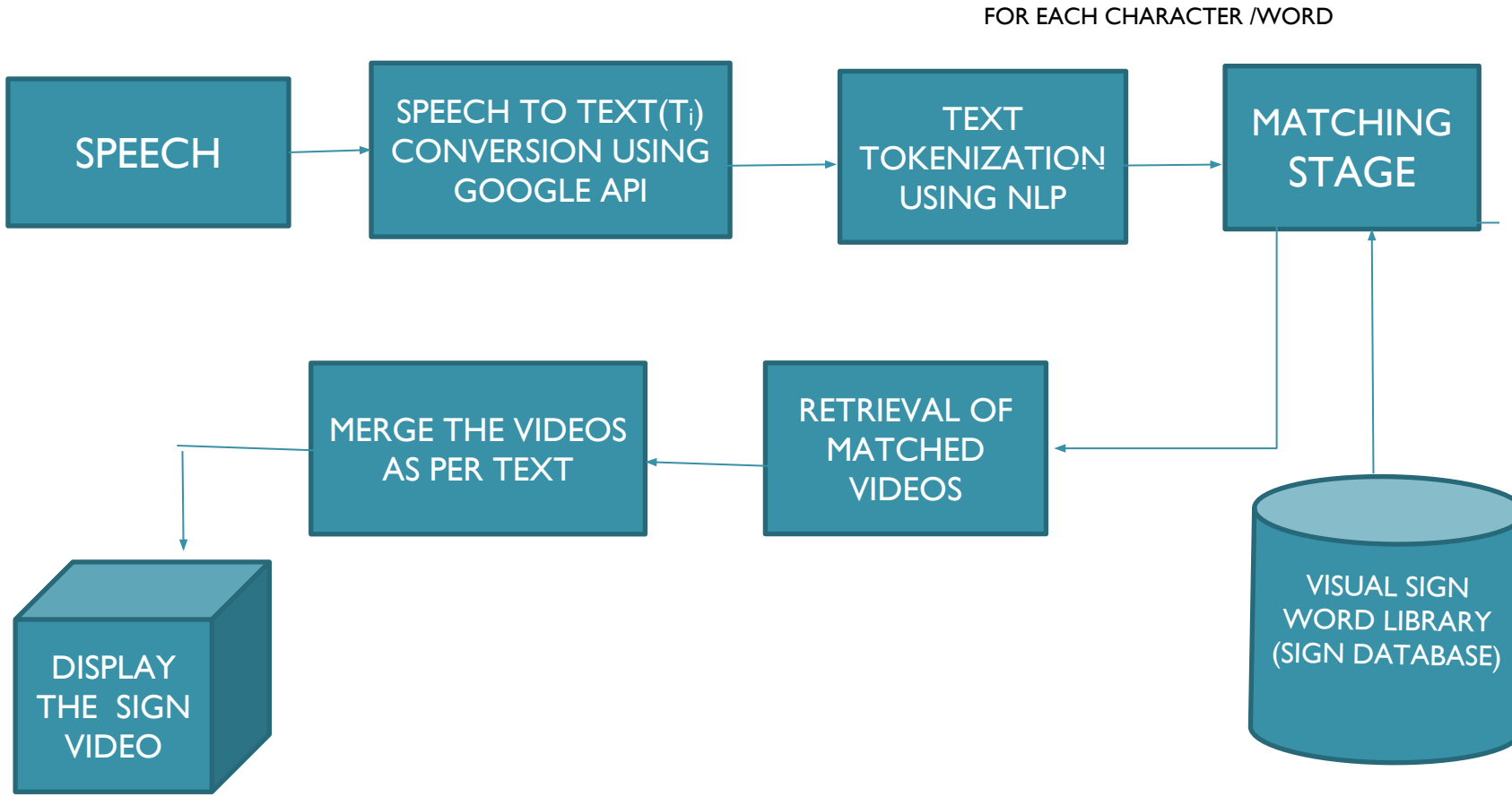
Introduction and Objective

- Sign language is a visual language that is used by deaf and dumb people as their mother tongue.
- With very less number of people understanding the sign language, communication seems really difficult for the deaf in every walk of their lives.
- We therefore introduce a web application and an android application that helps them communicate both within and outside their groups.
- This pilot study helps in the flow of conversation between the deaf and normal people as well as two or more deaf people.

Related Work :



Proposed Methodology

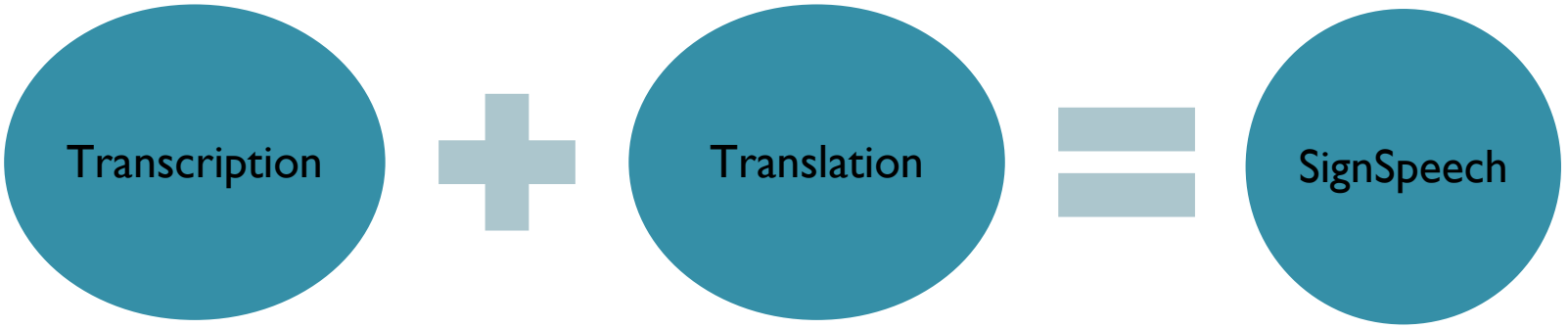


Implementation

SignSpeech functions in the following two stages:

- Speech to Text
- Text to Sign Language Video

Thus, it is based on the principles of:



Transcription

Turning Spoken language into written language

Challenge: Spoken language much looser in terms of grammar and syntax

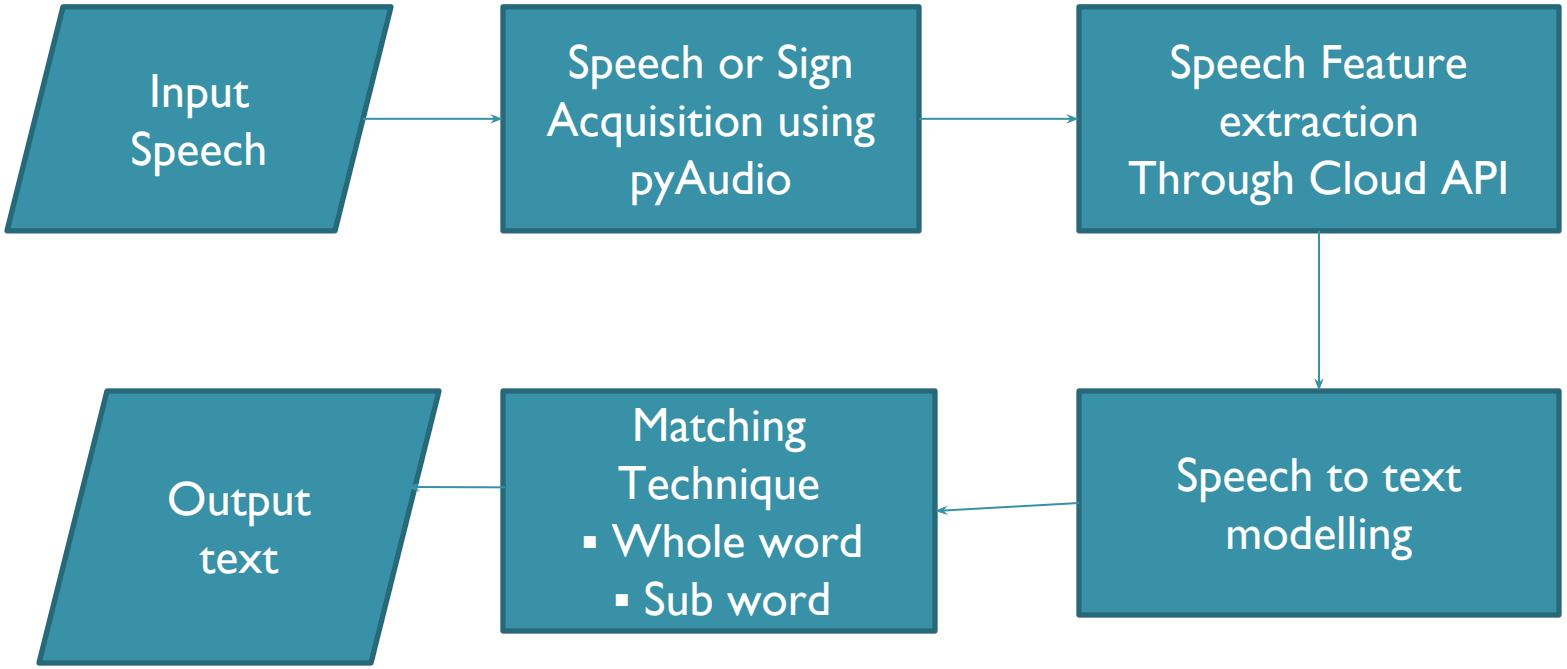
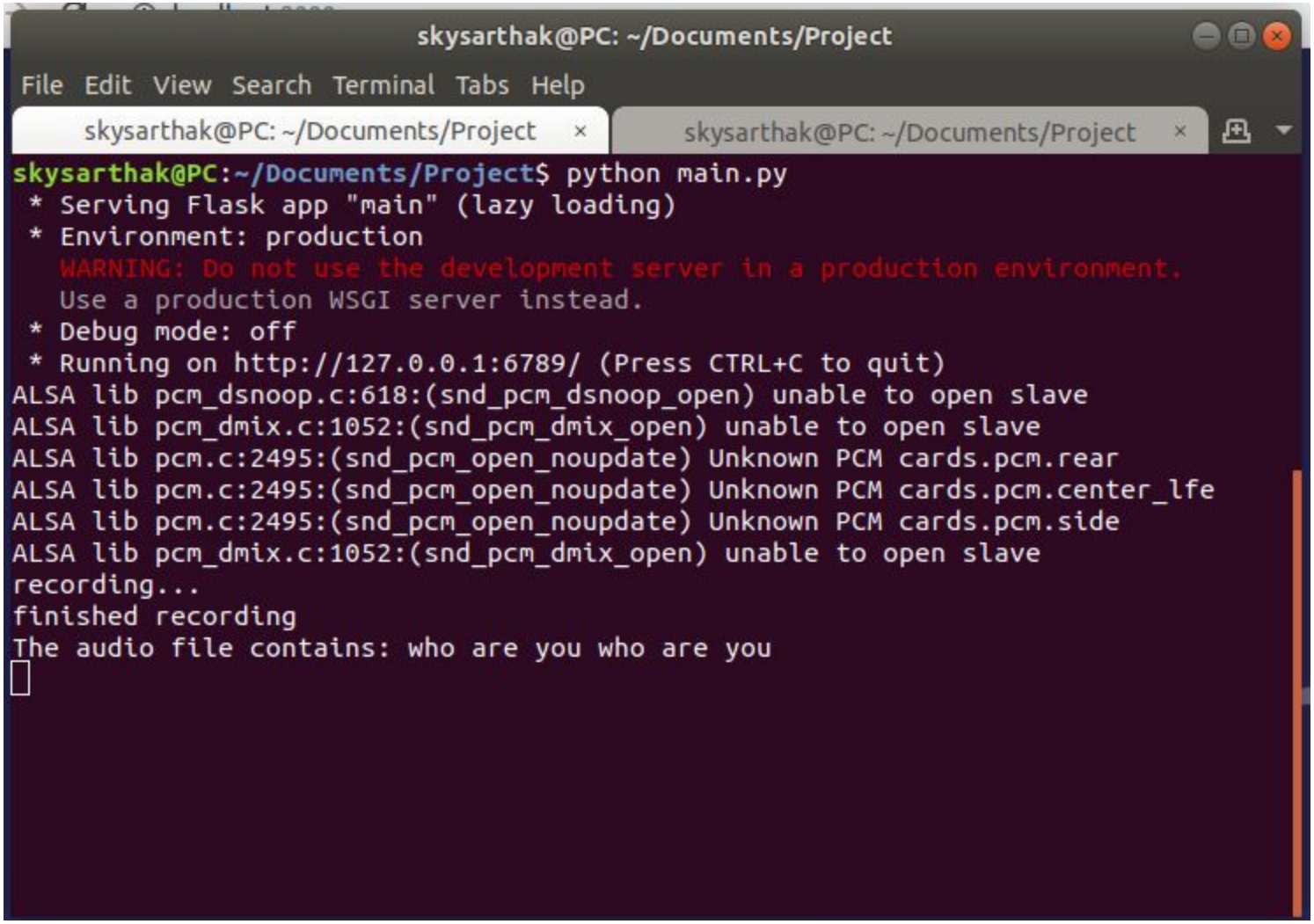


Fig. Stage-1 Speech to text

Transcription in SignSpeech



Translation

Turning the source text from the audio into a target language (i.e. Sign language in this case) that is well understood by the people involved in conversation.

The second stage of SignSpeech thus involves the following steps:

- Creation of Sign Language Video Library(database)
- Text Preprocessing
- Text Matching-Phonetic Matching using Soundex

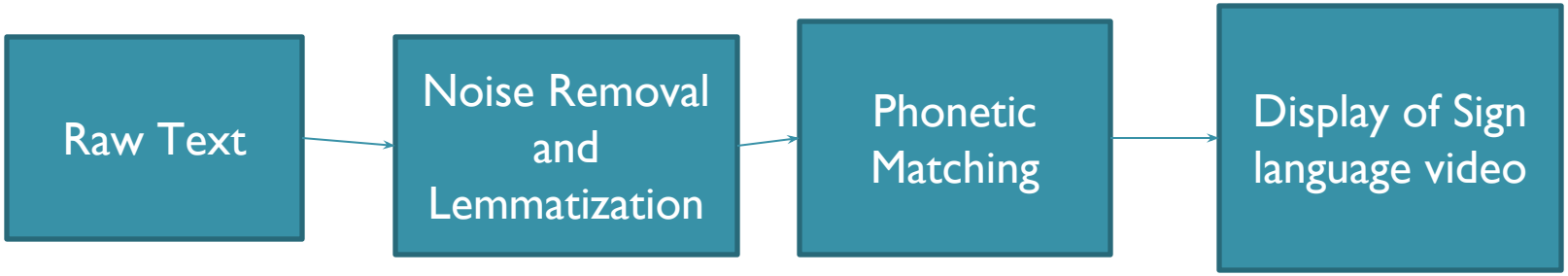
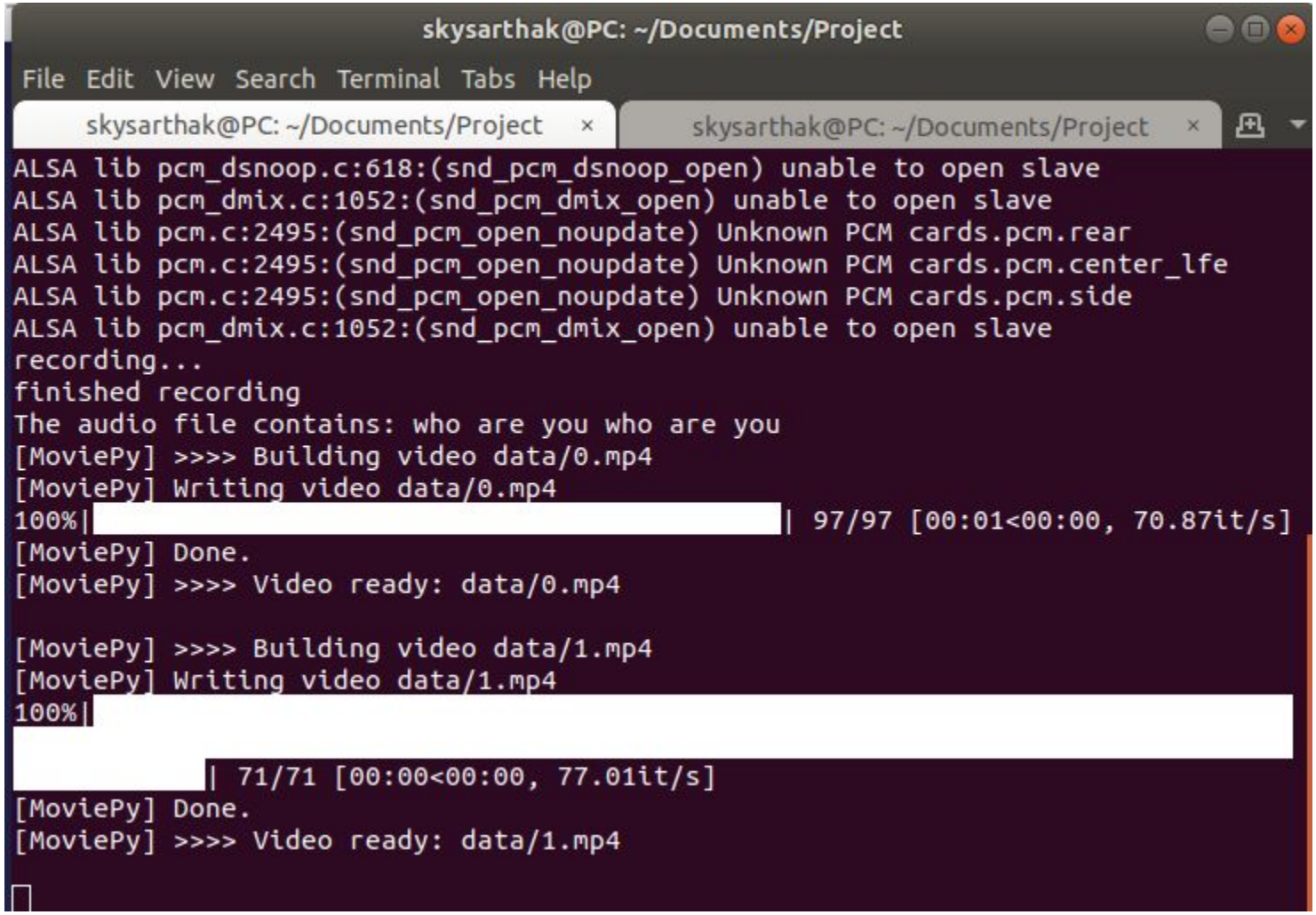


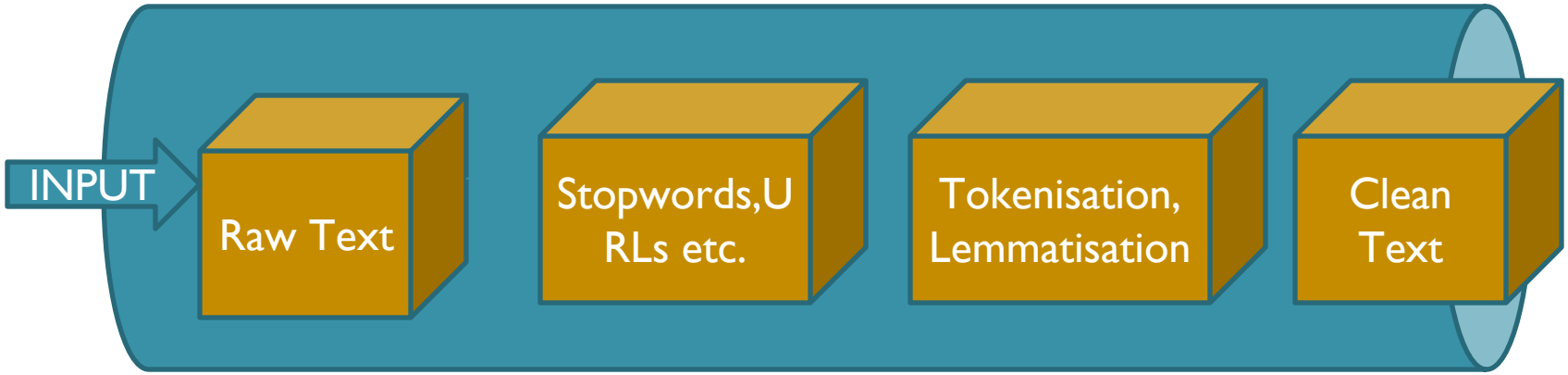
Fig. Stage-2 Text to Sign Language

Collection of Dataset-Sign Database



Noise Removal and Lemmatization

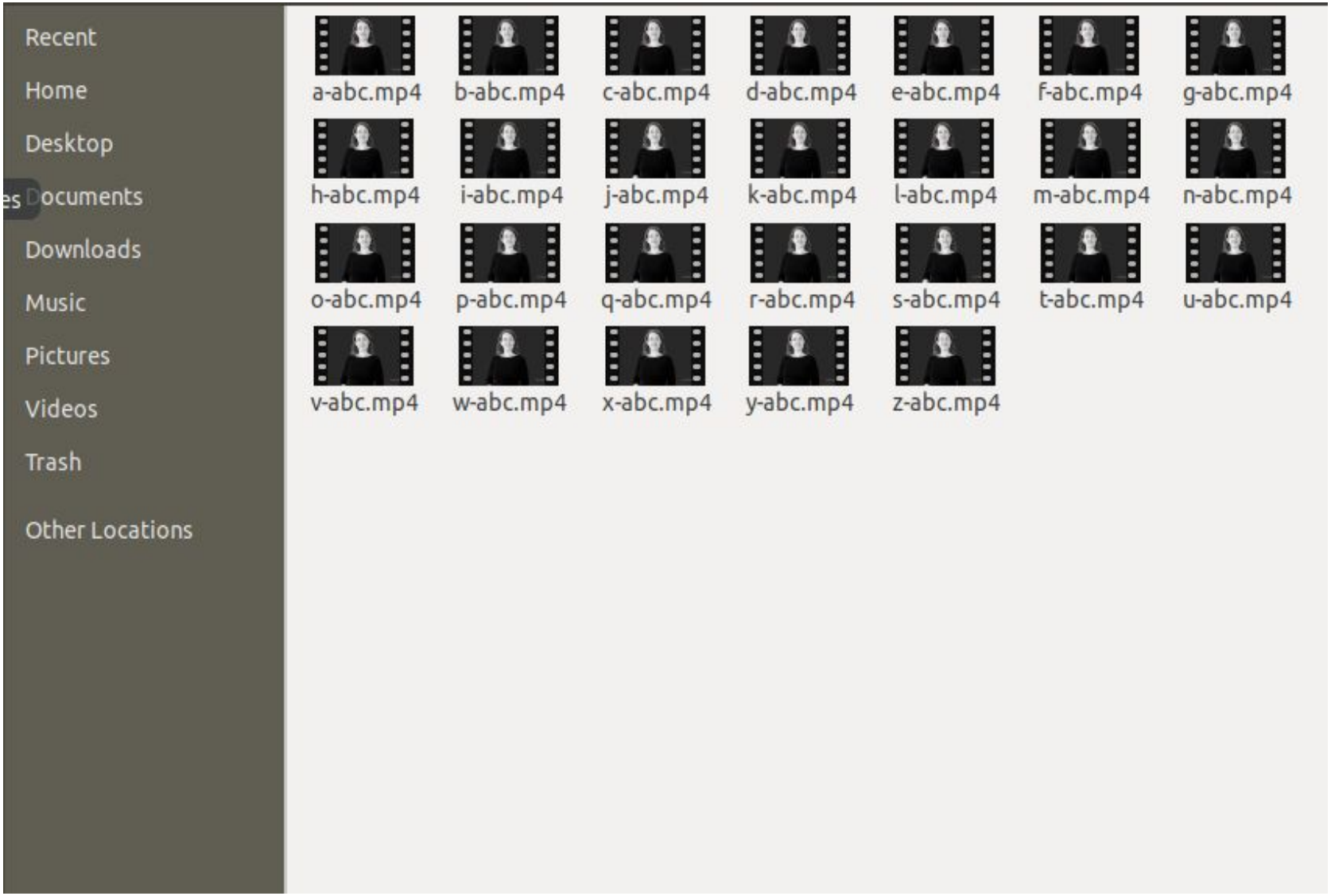
This forms the architecture for the text-processing pipeline as :



- Any piece of text which is not relevant to the context of the data and the end-output can be specified as the noise.
- Lemmatization, on the other hand, is an organized & step by step procedure of obtaining the root form of the word, it makes use of vocabulary (dictionary importance of words) and morphological analysis (word structure and grammar relations).

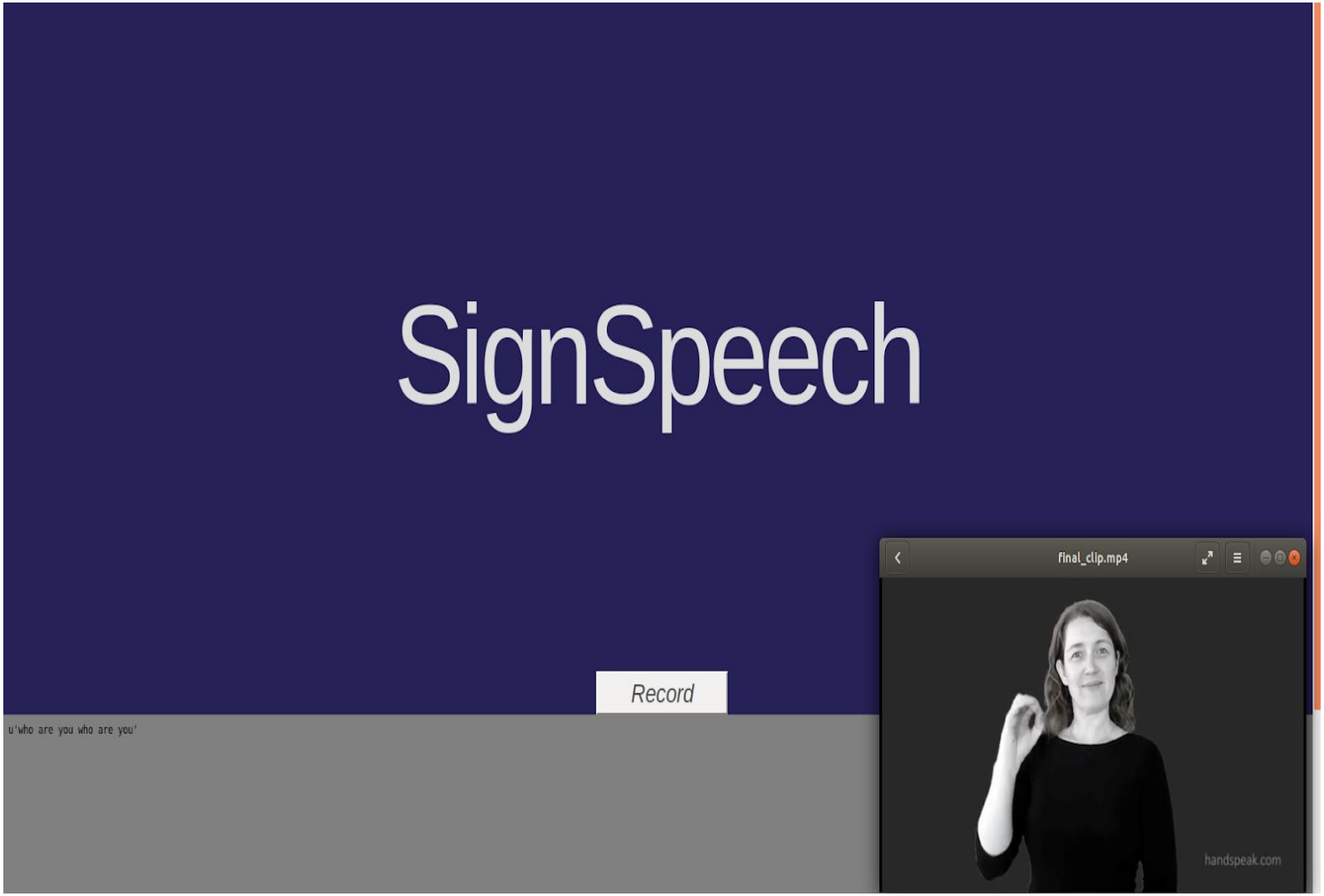
Text Matching

- A Phonetic matching algorithm (Soundex) takes a keyword as input (person's name, location name etc) and produces a character string that identifies a set of words that are (roughly) phonetically similar.
- The purpose of phonetic matching is to increase recall—to spread the net wide enough to catch any spoken text that might possibly match.
- useful for searching large text corpuses, correcting spelling errors and matching relevant names


Recent
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locations
a-abc.mp4
b-abc.mp4
c-abc.mp4
d-abc.mp4
e-abc.mp4
f-abc.mp4
g-abc.mp4
h-abc.mp4
i-abc.mp4
j-abc.mp4
k-abc.mp4
l-abc.mp4
m-abc.mp4
n-abc.mp4
o-abc.mp4
p-abc.mp4
q-abc.mp4
r-abc.mp4
s-abc.mp4
t-abc.mp4
u-abc.mp4
v-abc.mp4
w-abc.mp4
x-abc.mp4
y-abc.mp4
z-abc.mp4

Results

Results (Contd.)



Video Demonstration

Now, let us have a video demonstration of the simulation done by us. We have already discussed about the algorithms used in the implementation of SignSpeech.

This simulation gives an insight over the entire procedure and the obtained results.

Mode of Operations

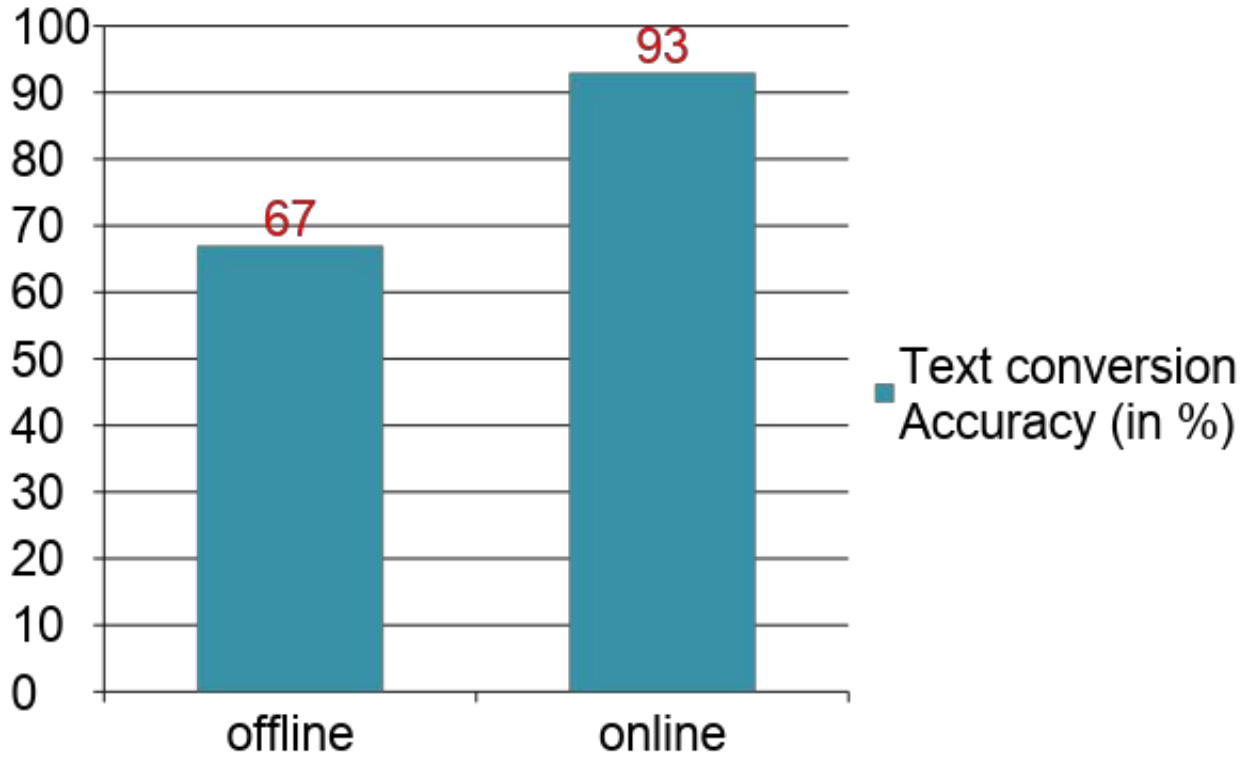
- To ensure that our model come up to the standards of best state-of-the-art approaches, model is implemented to handle Offline Mode(using Pico Voice-Cheetah) and multiple languages (using Goslate).
- A Comparison between these two modes of operation is then made and results are obtained.

Performance Analysis

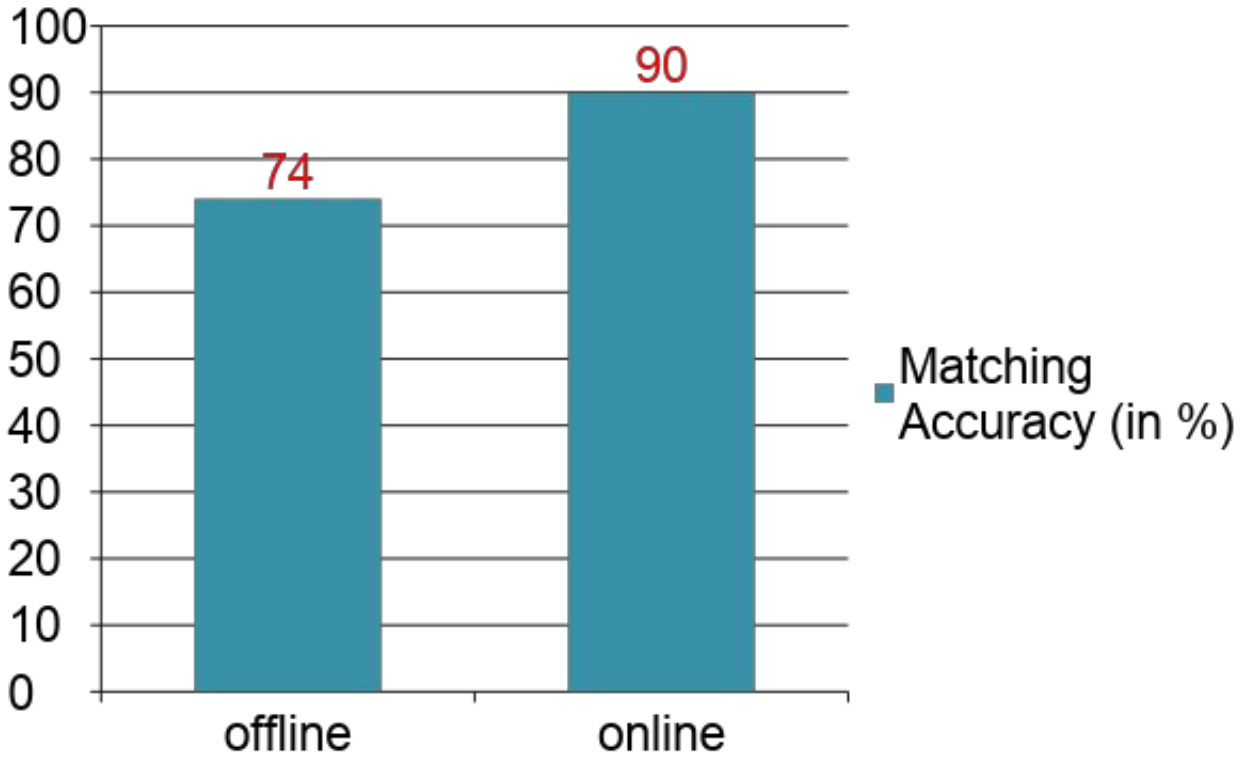
- The model is tested with 45 audio input samples and the accuracy percentage has been obtained using the formula:

$$Accuracy = \frac{Number\ of\ Correct\ Operations}{Total\ Number\ of\ Operations}$$

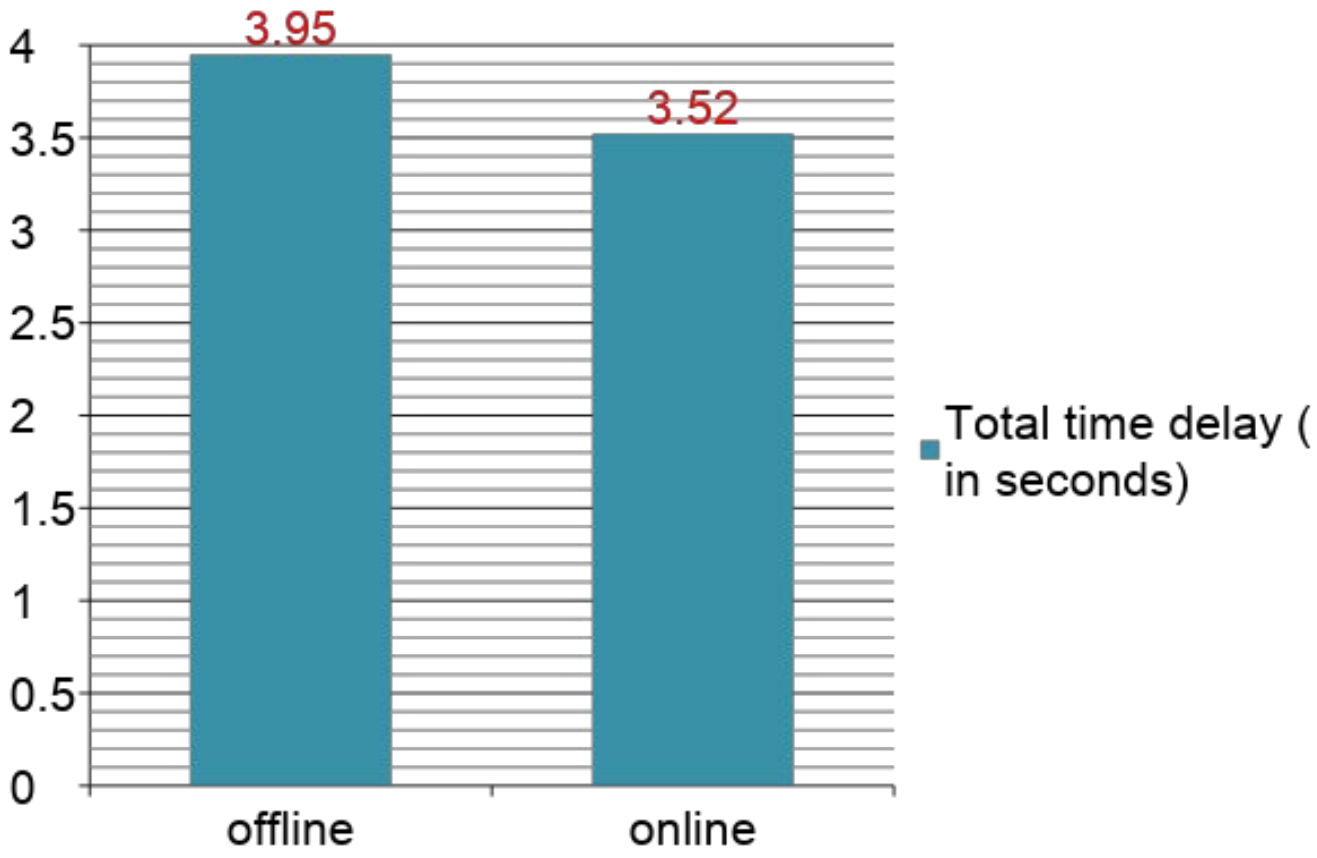
Comparison Results Analysis



Contd.


100
90
80
70
60
50
40
30
20
10
0
74
90
offline
online
Matching Accuracy (in %)

Contd.


3.95
3.52
4
3.5
3
2.5
2
1.5
1
0.5
0
offline
online
Total time delay (in seconds)

Future Scope

✔ modify SignSpeech to cover Sign language to speech conversion for the linguistic growth of dumb people.

✔ Incorporating non-hand signals like facial expressions and body language for complete understanding of the context.

✔ Integrating Computer Vision for 2-way Communication System.

THANK YOU