

## 5. CONCLUSION AND A ROAD-MAP

---

This chapter summarizes the research work into two parts. The first section gives the inferences that have been derived through this work based on the experiments and their results' analysis. From the experiences and the inferences through this work, the second section throws light on some of the areas where further work can be done.

### 5.1 THE CONCLUSIONS

After careful experiments and detailed analysis of the proposed algorithm, the following statements can be made as concluding remarks.

For 15 frames dataset, models need more epochs to recognize alphabets whereas for 20 frames, models need a smaller number of epochs to recognize alphabets for each class. This is because each epoch exposes a small subset of information of overall data when frame size is small. With fewer frames, the model might struggle initially to learn different parameters like speaker tone, lip movement, accent etc.

Among the five classes, Palatal and Labial classes are easy to recognize, as lips have crucial role in utterance of alphabets of these classes. Retroflex and Dental have low recognition rate as tongue and teeth are used respectively for the utterance of these classes. Dental class have better accuracy than Retroflex as teeth and tongue booth are involved in articulation of dental class alphabets, while in retroflex only tongue is involved. Guttural class has low recognition rate because these alphabets are spoken from throat, so it is difficult to differentiate between guttural, retroflex and dental when speaker's articulation is not proper.

Gujarati alphabets have limited recognition rate because of

1. Many alphabets have similar types of articulation
2. Alphabet recognition is difficult compared to words and sentences
3. People speak Gujarati but are not aware about proper articulation and pronunciation of Gujarati alphabets

In the GJVarna Dataset, each speaker has different time duration for alphabet articulation. Variation in utterance time affects the frame selection. Dataset is created for 34 consonants of Gujarati alphabet; hence, the number of speakers is limited. Each class has data of three consonants instead of a single consonant, which may also lead to low accuracy.

## **5.2 A ROAD-MAP TO FUTURE WORK**

Since the work for Gujarati language has been carried out for the first time, it has many future applications to be done. The dataset can be prepared with a greater number of trained speakers with fixed utterance time duration. Dataset for words, phrases and sentences can be prepared for next level work. Frame removal algorithm can be revised to remove starting and ending similar and closed lip frames.