CHAPTER 7

Summary and Future Scope

7.1 Summary of the present work

The necessity of understanding human language by electronic devices was felt for a long time. Specially, for the computer illiterate mass, who do not have adequate knowledge of extracting required information by themselves. The problem is catalyzed by the fact that till today the human computer interfacing is dominated largely by English. This significantly retards desirable penetration of Information Technology (IT) upto the expected depth. Hench, research and development in speech technology and development of Natural Language Processing (NLP) becomes a crying need for India.

India is a country having rich set of languages. Specially, the north-eastern part of India, which comprises of seven states, has a large number of languages and dialects spoken by the inhabitants. In Assam, **Assamese**, **Bodo**, **Rabha**, **Hajong**, **Mech**, **Lalung**, **Tipra**, **Chutia**, **Moran** are the main linguistic communities. Besides Assamese, the Bodo Language has got special constitutional status in the year 2003 and become one of the 22 scheduled languages of India. The Bodo language is one of the derivatives of Sino-Tibetan language family [52]. It is a tonal language, which has four dialects [53]. In case of Tonal Languages the effect of prosody is not very prominent [75]. But due to the socio-cultural influence of some non-tonal languages, some time the prosodic features can be prominently visible in case of tonal languages as well. This is the case in case of Bodo Language. Due to

the influence of Assamese, Bengali, Hindi, Bodo Language has significant prosodic variation.

The work undertaken in the present study has been summarized below along with observations and findings:

Chapter 1 is the introductory part of the thesis, where an introduction to the speech research has been given. The objective and formulation of the present study is also made here. Production of speech by a human being is a very complex phenomenon, which involves the interaction of a set of organs, namely **Nasal cavity, Hard palate, Alveolar ridge, Soft palate (Velum), Tip of the tongue (Apex), Dorsum, Uvula, Radix, Pharynx, Epiglottis, False vocal cords, Vocal cords, Larynx, Esophagus**, and **Trachea**. A mathematical formulation of the speech production system using two tube model and three tube model for vocal tract is explained in this chapter. Some commonly used terms used in Natural Language Processing (NLP) has been explained. The chapter ends with the application areas, where the outcome of successful speech research can be utilized.

The speech synthesis process has been discussed in chapter 2. The different speech recognition approaches like **Dynamic Time Wrapping (DTW), Hidden Markov Model (HMM), Artificial Neural Network (ANN) are discussed with their relative merits and demerits.** A discussion on the history of Speech Recognition Research starting from 1950's till date along with the tools available for research during those time is also discussed in this chapter.

A brief history of Bodos and Bodo language has been made in Chapter 3. The general structure of Bodo language and its hierarchical construction has been discussed here. The historical background of Bodo community and the origin of the language spoken by them is clearly explained. It also includes a description of Bodo vowels and consonants and their occurrence characteristics at different tongue position. The chapter ends up with some discussion on the Tonal characteristics, word order, intonation and prosody with respect to Bodo Language.

A study on the cepstral characteristics of the vowels and words of type VC, CV and CVC type have been made in **chapter 4**. It has been observed that LPC based cepstral coefficient (LPCC) do not play significant role in the identification of male and female informants of Bodos. But Mel-frequency Cepstral Coefficient (MFCC) shows distinctive features for male and female informants through the utterances of vowels and words of types CV, VC, CVC.

It has been observed that in case of CV type words, with high tone, there is a remarkable differences in MFCC with respect to male and female informants. At the same time, with a low tone, these differences are not distinct. Similar characteristics are also observed in case of VC type of words. When the words are of CVC type, the MFCC is almost stable in case of male informants and with a slight variation in case of female informants. Thus, the MFCC features obtained through the CV and CVC type words utterances may be a useful parameter for sex identification and verification for Bodo native speakers.

The formant frequency measure and analysis of Bodo vowels and words of **CV, VC** and **CVC** type has been made in **chapter 5**. It has been observed that **Third Formant (F3)** does not play any significant role for the identification of speech and speaker. The **First Formant (F1) and second formant (F2)** frequency variations are found distinct for all vowels. It is observed that the Second Formant (F2) of **/a/, /e/, /o/, /w/** shows very distinctive features with gender variations.

As far as CV type words are concerned, change in the Formant Frequency, especially F1 is gradual with respect to female informants. On the other hand, the change of F1, F2 and F3 all are gradual in case of male informants. Thus, through the tonal feature of CV type words it is not a preferable way to identify the speakers.

In VC type words, there is a marked departure of F1 from F2 and F3 depending on tones. Similar results were obtained for both male and female informants. Further, it has been observed that the change in frequency in case of F2 is abrupt for low tone utterances. Thus, the tone feature of VC could be a measure for further conformity for sex identification.

In case of CVC type words, both in case of Male and Female informants, for words with high tone, the F2 play an important role in speaker verification.

Prosody is very important to give naturalness to any TTS system. It refers to both pitch which consists of intensity and fundamental frequency (F0) and to the placement of pause also known as duration in speech. It is the key element for

making synthetic speech sound natural and acceptable to native listener. From the prosody of the uttered speech, one can identify gender, age and even individual. Prosodic features are expressed by different prosodic component in different types of language. In English, prosodic information is carried by **lexical stress**, in case of languages like Japanese, it is the **accent** and in case of tonal languages, like Bodo it is the **tone.**

In general, speech prosody can give information about the speaker's gender, age, emotion regardless of what is spoken. Characteristics of prosody can be well analyzed by using the variation of pitch within the word or sentence. The **chapter 6** highlights different methods used for pitch detection, among which **Auto Correlation Function (ACF)** and **Average Magnitude Difference Function (AMDF)** are elaborated and used as a tool for the current study. The chapter discusses about the tone variation about **monosyllable, disyllable** and **trisyllables** and results have been analyzed graphically. The results obtained through the present study are very interesting and found promising for speaker/speech verification in a robust environment.

As Bodo language expresses the prosodic feature by use of Tone, so it is very important to study acoustically the tonal characteristics of the language. The prosodic feature of Bodo Language is studied in terms of Mel-Frequency Cepstral Coefficient (MFCC), Formant Frequency namely F1, F2 and F3 and Pitch variation. All the three tools used supported and advocated their relative utilities for better and robust speech, speaker recognizer.

7.2 Future Scope

There is enough scope of extending the current study. The present study gives a constructive platform to continue further research. Taking the present study as a basic platform, there is enough provisions for incorporating new ideas, techniques for further improvement and enhancement of the present work. The study on prosodic feature of Bodo language could be further enhanced considering more different types of words like CVVC, VCCV etc. Such study would help for better realization of the prosodic features of Bodo language through varied application platforms like –TTS, development of G2P rules and syllabification rules etc. The TTS (Text To Speech) converter may become more natural and intelligible after successful integration of prosodic features to it. With all the cited features the present model could be used as a role model for integrated study of other languages of this region.