CHAPTER 8

SUMMARY AND FUTURE SCOPE

8.1 Introduction: Summary

Human beings communicate with one another with the help of spoken language i.e Speech as well as body language, facial expression etc. Emotional information contained in the speech is an important part of speech. The same set of words when spoken in different expressions conveys different meanings. Every person has a different style of expressing their emotions. Every speech contains some amount of emotion. The expressions / emotions that are infused into the speech of a person speak a lot about the emotional state of that person. Information which is not spoken in words is also conveyed through the emotions contained in the speech. Different types of expressions are conveyed by changing the acoustic signal by using different 'tones of voice' instead of changing the words which are being used or the arrangement of the words in the utterance. The knowledge of when, how and where prosody or emotional contents / information are infused into the speech signal is very important. This is because this helps to design and develop a HMM-based Emotional Speech Synthesizer / recognizer.

The whole process consists of a planning stage followed by an implementing stage. The implementation stage is of utmost importance and has to be monitored closely. This is achieved by the analysis of acoustic and prosodic features obtained from the speech signals under consideration.

The Hidden Markov model was basically chosen as the implementing model because it is considered capable of computing and implementing the system. It can run on PDA and computers easily. The entire model is descriptive of the characteristics of human speech with emotion mainly based on the experimental observations. The HMM probability analysis can be predicted with the help of formulae's. An acoustic speech signal incorporating emotion can be predicted by choosing from some in-build models for prosody. But, the main hindrance for the design of such systems is that there are many problems at various levels of implementation which needs to be programmed in detail. A considerable amount of knowledge has to be gathered about how human speech production system works and the constraints which are involved with it. Hence, to solve this problem, it is essential to design a model which characterizes the human speech. The computational model which is used to design HMM based synthesizer incorporates the properties of human speech which are dealt with in the different chapters of the thesis. The model should be capable to incorporate into a high-level speech synthesizer. In recent times designing of man-computer interactive devices is a major goal of many research works. But, this is mainly achieved with the help of the English Language. Hence, the advancements of IT are beyond the reach of all the people who are not familiar with the English Language. This is a big drawback for a country like India, which has many varieties of languages and its dialects in all parts of the country. The achievements in the field of speech should be implemented with all the recognized language of India and this requires immediate attention. Achieving this objective will help bring the benefit of the Information Technology to every person's doorstep and even those who do not have any formal education.

A brief discussion of the different chapters of the thesis is as follows:

(i) CHAPTER 1: Basics of Speech Research

This chapter of the thesis deals with the importance of Speech research. It discusses the process of Speech Production process. It gives stress on the importance of man-machine communication for the design and application of modern systems. It also gives an introduction to the speech synthesis process and the different approaches of the speech synthesis process. This chapter gives an introduction to the analysis, synthesis and recognition of the speech production process. It discussed the problem Statement and the necessity of the proposed Research.

(ii) CHAPTER 2: Review of Literature

This Chapter discussed the methods used for the Speech Synthesis Process. It also discusses the historical developments of the Speech Synthesis process. It also focussed on the achievements in Speech Technology from 1960’s to the recent times. It also deals with the evaluation of Text-To-Speech Systems. The proposed approach has been discussed in this chapter.

(iii) CHAPTER 3 : Research Methodology

This method adopted for doing the present research work is discussed in this chapter. The process of Bodo Emotional Text selection, speaker selection, setup used for recording as well as the recording are discussed here. The methodology and the tools that are used I the present research work is also discussed in this chapter.

(iv) CHAPTER 4: An Introduction to the Bodo Language

This chapter gives a brief history of the Bodo’s and the Bodo language. The linguistic structure, historical background, origin as well as its hierarchy is discussed. It also takes into account the sound variants of the Bodo Vowels together with the characteristics of utterance in different tongue positions. Also discussed here are the Bodo script, dialects of the Bodo language, Bodo dialect region, present Bodo dialect area, lexical variation, phonological variation and grammatical variation in the Bodo language. It also highlights the structure of the Bodo language like phonological structure, vowels, consonants, diphthongs, juncture, Bodo tones, morphological structure and syntactic structure of Bodo language. The Tone system and the intonation pattern of the Bodo language are also discussed in this chapter.

(v) CHAPTER 5: Estimation of Formant for Bodo Vowels and Bodo Words of Different Types

This chapter discussed the importance of the study of formant frequency measures. The analysis of different Bodo vowels and words of different structures like CV, VC and CVC type are made with respect to their formant frequencies. During the analysis, some characteristic features of these are obtained. It has been found that the Formant Frequency F2 plays a significant part in the identification of emotional content or emotional status of Bodo informants. The characteristics features of CV, CVC and VC of words were also extracted and the result were analysed for their emotional/prosodic features. The observations helped in the synthesis and the recognition of the Bodo Speech in the proposed model.

(vi) CHAPTER 6 : Acoustic Representation and Prosodic study of Bodo Vowels and different types of Bodo Words: The Acoustic and Prosodic features of Bodo vowels and words of CV, CVC and VC type are under consideration in this chapter. Each voice signal is divided into frames each containing 250 samples and corresponding MFCC, ZCR and STE has been estimated. It is realized that these features, both acoustic and prosodic have a very important role to play for analyzing the difference between the different emotional states of both speakers- male as well as female. A distinctive variation is seen for all the different emotional state for both male and female informants.

(vii) CHAPTER 7 : Analysis, Synthesis and Recognition of Emotional Bodo words with Reference to its Prosodic Features using Hidden Markov Model (HMM):

The use of HMM for speech recognition and synthesis and its use in recent times is described in this chapter. It also gives an overview of the structure of HMM. It also explains the training process of the HMM's for the synthesis and recognition of Bodo emotional speech. Here, the HMM feature are extracted and comparison is done for the three different emotions i.e. neutral, surprise and angry modes. Also investigated has been done for words of different structures spoken in various emotional moods. A comparative study is done on the values which have been generated by the use of different techniques like MFCCs, HMM, etc. It is seen that these kinds of systems perform efficiently in suitable environments. But the quality degrades in the presence of noise. This is because the properties of speech are based on the acoustical and prosodical features of speech. HMMS are used and trained to represent these statistical properties. But, when noise is present in speech, HMMS are also trained to describe noise which leads to the degradation of the system.

8.2 Future Scope of the Present Research Work

Over the past few decades as well as in recent times, significant advancements have been made in the field of Speech synthesis and recognition. Easy and user friendly methods for interacting with computers is the need of the present times. The main focus is to make technology more approachable and useable to human beings who are not capable of handling computers .The present research work primarily focuses on the issues relating to emotional speech synthesis and recognition using HMM. The model developed for emotional speech synthesis / recognition process distinguishes between two stages, with each stage focusing on the utterance plan of the speaker. The prime objective of the HMM model is :

i) It is mainly used as a computational model.

ii) HMMs help the system to extract the phonetic information with emotional content.

The HMM-based model is thus built up to integrate tightly the prosodic (emotional) as well as supra-segmental features of speech containing emotions in a methodical and efficient way. Prosody gives us knowledge about the phonology of the segments. It considers and determines all the aspects of speech production which contain emotions.

The focus is on the use of various kinds of data structures used at various levels, which is important for building successful computational modeling. This dynamic approach will allow data manipulation any kind i.e., any new or old speech sound. It is also important that the model occupies very small space, only then it will be possible to use it in small portable device applications.

In the present study, emphasis is also on the dynamic pattern of speech. The study of prosody also helps to analyze the dynamic nature of speech. The part of the speech signal containing the prosodic features is classified hierarchically. The lowest level of the hierarchy are the sub-domains which include sub-syllables and syllables along with lexical objects. These are known as the framework of utterances. Instantiation of a particular words utterance in various emotions is mainly done within the hierarchy of the organised prosodic framework.

The scope of the present research work is predominantly concerned with the struggle associated with emotional speech synthesis / recognition of the Bodo language. This research work on “BODO EMOTIONAL SPEECH SYNTHESIS AND RECOGNITION USING HMM” is an important step towards the investigation of the acoustic / prosodic features of the Bodo language on a scientific platform. In this study, the features of speech that has been considered are mainly Format Frequencies (F1, F2 and F3), Mel Frequecy Cepstral coefficients, Short Time Energy and Zero Crossing Rate, etc. The desired results thus achieved depend on recordings of emotional speech of 30 speakers both male and female. These are analyzed on the Bodo vowels, and different type of words of Bodo CV, VC and CVC type which are spoken in different emotions. In the present research work, highly expert software and hardware system is required. The resolution and authenticity of the recognized / synthesized speech also requires a lot of improvement. This can be achieved by applying Fuzzy logic, Artificial Neural Networks etc. Moreover, the total number of speakers and words can be increased and thereby the corresponding acoustic and prosodic characteristics can be chosen as parameters for any language speech recognized / synthesized process. So, by increasing the number of words in the database covering more emotional states from a larger number of speakers from different districts of Assam and its adjoining areas, a better quality system can be built. The results reflected in the chapters will give ideas and encouragement for further investigation of the Bodo language as well as other Indian languages with a more scientific approach.