

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Methodology: An Introduction

The systematic work undertaken to increase the knowledge stock, which may include the knowledge of culture, society, humanity, science etc is known as Research Methodology. It is the formal work undertaken to make new developments. It may also be defined as the way to systematically solve any kind of Research problem [30]. It explored the various steps adopted in analyzing the research problem and developing the logic for solving the problem. The development of the methodology is the most important work for a researcher. The method undertaken in two different problems may be same but the methodology may not be the same. Considering the above points, in the present research work the following issues have been dealt with.

3.2 Emotional Speech Corpus Development

The emotional speech corpus is an important part of the emotional speech production system. The quality of the synthesized speech is mainly dependent on the quality of its acoustic units. There are several factors like the speech corpus in use, from where the speech units have been extracted, the accuracy of labelling, the prosodic richness of speech units etc, that are of importance. The process of preparing the speech corpus mainly involves the following steps:

- (I) Pre-processing of the collected text.
- (II) The selection of emotionally rich sentences.

(III) Recording of the emotionally rich sentences by both male and female speakers.

(IV) Orthographic annotation of sentences.

3.2.1 Bodo Text Selection and its Normalization

The Bodo sentences are selected from various sources like story books, lectures by experts and academicians on various issues like culture, education, weather, sports, economy, science etc. Sentences having high emotional content are chosen for the present study. The sentences are continually checked for their uniqueness and other special characters. The sentences, when uttered, must possess different intonation and annotation. In the present study the Bodo text are stored in UTF-8 format for the recording and further processing.

3.2.2 Bodo Emotional Speech Recording, Duration and Format

The software and recording format used in the present study:

- I. Softwares used for recording: Cool Edit Pro, Wave Surfer 1.8.8, Audacity 1.3.6.
- II. Resolution: 16 bit PCM, Mono.
- III. Sampling Frequency used: 1600Hz
- IV. Duration of recording: 5 hours
- V. Text format for recording: UTF-8

3.2.3 Setup used for Recording

- I. The Microphone used for recording was Behringer C-IU, USB studio condenser, USB studio condenser, cardioid pickup pattern.
- II. There was a minimum distance of about 8 inches between the microphone and the speaker.
- III. The recording was done in a noise proof environment.

3.2.4 Speaker Selection Process

The criteria that have been followed when selecting the male and female speakers are as follows:

- (a) The age of the speakers was between 20 to 40 years.
- (b) Speaker's mother tongue: Bodo.
- (c) Qualification of speakers: Graduate / Post Graduate

Phonetic experts of the Bodo language supervised the recording procedure so as to ensure the accuracy and consistency of the recorded speech both phonetically and prosodically.

3.2.5 Procedure for recording

The following criteria have been followed while recording the selected text:

- a) The TTS lab meant for acoustic study of speech constructed in the Department of Instrumentation and USIC of Gauhati University was used to do all the recordings.
- b) Two sets of the recording materials were printed, one for the speakers and one for the supervisors.
- c) The phonetic words, embedded in the bodo sentences are extracted from the sentences and stored. A similar method is followed for sentences also. The speakers are provided a reset time of around 10 to 15 minutes after each session of recording for a particular item. The maximum duration that was allowed for continuous recording was half an hour.

3.2.6 Creation of the emotional database

- (i) The recorded sentences which contain the emotions are stored in a separate directory if the form of a wave.

(ii) The file that has been created to store the wave is kept as follows:

“text1(m1/f1)_yyy.wav

(iii) The corresponding text of the wav file is saved in the same directory. The text file is stored in UTF-8 format

“text1(m1/f1)_yyy.text

3.3 Methodology used for analyzing MFCC and Formants

The recorded speech data is used for analyzing the Mel frequency Cepstral Coefficients (MFCC) and the formants like the First(F1), Second(F2) and Third(F3) formant frequencies for :

(i) Bodo Vowels

(ii) CV, VC and CVC structure for selected words, C:Consonant , V:Vowel

The separation of bodo vowels, CV, VC and CVC structures from the recorded sentences and phrases are carried out by using **WaveSurfer 1.8.8**. The speech analyzing tools i.e **MATLAB 14a** and **CochLEA Toolbox** are used to do the analysis of the speech properties.

For analyzing the MFCC, a set of MFCC Coefficients is computed for a speech frame of around 20ms. In the present research work, 20 MFCCs has been computed for the utterances of Bodo vowels, and different words of CV, VC and CVC types.

In the Formant Frequency analysis, the Formant Frequencies (F1, F2 and F3) are computed and analyzed. Each digitized voice that is uttered is divided into 50 frames, each of duration 20 ms. Each frame consisted of 441 samples for which the formant frequencies are computed and analyzed.

3.4 Tools used for Recording and Synthesis of Speech

3.4.1 Audacity

It is an open source software, which is free and used for digital audio recording. It can be used in platforms like windows, Linus, Mac OSX, etc. Audacity was first started by Rogu Dannenberg in May 2000 at the Carnegie Million University. The Audacity software can also be used for the post-processing of any type of audio which may include podcasts by adding effects like trimming, normalization, recording audio from various multiple sources. It is also useful for recording and mixing many different sounds. The main features of the Audacity software are as follows:

- The import and export of wav, MP3, AIFF files.
- The record and playback of sounds.
- The editing of sounds like cutting, copying and pasting.
- The mixing of multiple tracts.

3.4.2 Wave Surfer

This is a software distributed by a free software license which is used for editing audio used for analysis of acoustic phonetics. It is a program which can be used for interactive display for spectral sections, sound pressure waveforms, pitch track transcriptions etc.

The basic features of this software are editing operations of audio such as copying of sounds, pasting them, zero-crossing adjustments, normalization, replacement with silence elements, DC-removal etc.

3.4.3 PRAAT

PRAAT is a free software that can be used for analysis of phonetics in speech [58]. It was developed in the university of Amsterdam by David Weenick and Paul Boersma. It can be used on various platform like windows, Linux, Unix etc. This software can also be used for synthesis of speech, analysis of speech etc. In the present research work, the software is used for recording analysis and the synthesis of the emotional speech sounds.

3.4.4. Cool Edit Pro

Cool edit Pro is a sound editing program that can be used in windows and is highly advanced [59]. The main features of this software are:

- Filtering sound by digital signal processing effect (DSPE).
- Multiple sound tract function(64 tract simultaneously)
- Ability to accept plug-ins by expanding its capability.
- The capability to create batch processing files.

This software can be used to import and export from a wide variety of formats like wav, MP3, WMA etc.

The emotionally rich speech sounds have been recorded by using cool edit pro 2.0. It is then digitized by using a sampling rate of about 16 Khz. The sound signals are then stored as .wav file in URF format. They are also stored separately for male and female according to their age groups. The platform that has been used is windows along with a high definition audio card as well as 16 bit PCM resolution. The voice recording is done using high quality head microphone. The correct pronunciation of words has been examined by experts in the Bodo language.

3.5 Selection of Sample

The samples used in the present study are vowels, and words of different structure in the Bodo language. Male and female speakers have been randomly picked up.

3.6 Filtering of noise from the speech data

A noise free environment has been used for recording the speech data. The speech signal has to be processed further for analyzing its properties. Hence, the digitization of the speech signal is done which is carried out and the sampling rate used is 16KHZ.