Final Report on

**Text to Speech using FreeTTS**

****

Submitted to

**Department of Computer Science and Engineering**

**Nepal Engineering College**

in Partial Fulfillment of the

Requirements for the Degree of B.E. in Computer

By

*Kiran Chaudhary (011-327)*

*Ashimita Ghimire (011-301)*

Supervised By: *Krishna Bikram Shah*

Submission Date: *01/01/2020*

ABSTRACT

This project report confer the tools and methodology used in developing a Nepali Text to Speech Synthesis System using FreeTTS and is written entirely in the Java programming language using FreeTTS synthesizer. Vocalized form of human communication is Speech. Here the Nepali Language is Synthetized based on formant approach and the use of most popular generic frameworks FreeTTS that is available in public domain for the development of a TTS synthesizer. The Text To Speech Architecture putting more emphasis on a components, namely Natural Language Processing (NLP) rather than Digital Signal Processing (DSP) component. Nepali language being mostly used language in Nepal and some parts of India and abroad, a text-to-speech (TTS) synthesizer for this language will prove to be a useful Information and communication technology (ICT) based system to aid to those majorities of people who are illiterate and also to those who are physical impairments like visually handicapped and vocally disabled physically handicapped. This ability to convert text to voice may reduce the dependency, frustration, and sense of helplessness of these people. The system can be extended to include more features such as more emotions, improved tokenization, interactive options and the use of minimal database.

**[Keywords**: *TTS, Prosody, Di-phone, Phoneme, Concatenation, Speech Synthesizer, Nepali vowel and consonants, Speech.]*

ACKNOWLEGDEMENT

We would like to extend our gratitude towards Department of Computer Engineering, Nepal Engineering College and Pokhara University for providing a platform, opportunity to undertake an exclusive Major project.

We would like to thank all the members in Department of Computer Engineering and Project Coordinator along with all the teachers for providing help in the every step during the development of the project.

Furthermore, we would like to give special thanks to our Project Supervisor Mr. Krishna B. Shah for providing knowledge, support, help and encouragement in accomplishing the task that are needed for the development of our Major project.

Thank you, everyone.

Ashmita Ghimire (011-302)

Kiran Kumar Chaudhary (011-311)

***nec***

April, 2016

Contents

[ABSTRACT i](#_Toc43371829)

[ACKNOWLEGDEMENT ii](#_Toc43371830)

[List of Abbreviations iv](#_Toc43371831)

[List of Figures v](#_Toc43371832)

[List of Tables vi](#_Toc43371833)

[Chapter 1 1](#_Toc43371834)

[Text To Speech 1](#_Toc43371835)

[1.1 Introduction 1](#_Toc43371836)

List of Abbreviations

|  |  |
| --- | --- |
| **Abbreviation** | **Explanation** |
| TTS | Text To Speech |
| OCR | Optical Character Recognition |
| NLP | Natural Language Processing |
| POS | Part of Speech |
| NeLRaLEC | Nepali Language Resources and Localized for Education and communication |
| ICT | Information and Communication Technology |
| WAV | Waveform |
| ISO | International Organization for Standardization |
| TD-PSOLA | Time Domain- Pitch Synchronous Overlap Address |
| MOS | Mean Opinion Score |

# List of Figures

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Title** | **Page No.** |
| Fig 1.1 | Schematic Representation of text to speech (TTS) system | 2 |
| Fig 1.2 | A simple functional diagram of TTS system | 3 |
| Fig 3.1 | System Overview of TTS | 17 |
| Fig 3.2 | Basic Flow Diagram of Nepali TTS synthesis System | 20 |
| Fig 4.1 | Flowchart of Nepali TTS | 24 |
| Fig 4.2 | DFD-level 0 | 25 |
| Fig 4.3 | Use Case Diagram | 26 |
| Fig 4.4 | UML Class Diagram | 27 |
| Fig 5.1 | Evaluation Chart of Vowel, Consonant, and Number | 32 |
| Fig 5.2 | Pie Chart Representation of Accuracy of Pronunciation | 33 |
| Fig 5.3 | Evaluation Chart of Words | 34 |
|  |  |  |

# List of Tables

|  |  |  |
| --- | --- | --- |
| **Table No.** | **Title** | **Page No.** |
| Table 4.1 | List of phone set | 22 |
| Table 5.1 | List of Numbers, Vowel, and Consonant used in test cases | 30 |
| Table 5.2 | Evaluation Table for Numeric Keywords | 31 |
| Table 5.3 | Evaluation Table for Vowel Keywords | 31 |
| Table 5.4 | Evaluation Table for Consonant Keywords | 32 |
| Table 5.5 | Evaluation Table for Individual Words | 33 |
| Table 5.6 | Weight of Words in each Categories | 34 |
| Table 5.7 | Evaluation of Total Words for Accuracy | 35 |

Chapter 1

Text To Speech

# 1.1 Introduction

Speech is the primary mode of communication between humans[1]. Speech Synthesis is an artificial production of the human speech. Speech Synthesis is the process, which allows transformation of the string of phonetic and prosodic symbol into a synthetic speech signal. TTS technology is the branch of artificial intelligence[2]. TTS system is a process through which input text is analyzed, processed, and understood and the text is rendered to digital audio and then spoken.

Text to speech (TTS) synthesis is the automated transformation of a text into speech that sounds as close as possible, as a native speaker or the language reading the text. Most Text To Speech Systems can be categorized by the method that they use to translate phonemes into audible sound. Some of them are Prerecorded, Formant, Concatenated, etc. The following table shows the comparison

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Prerecorded** | **Concatenated** | **Formant** |
| **Resource**  **requirement** | Very large storage, Small memory | Large storage, Very large memory | Low storage, relatively small memory |
| **Vocabulary** | Limited | Unlimited | Unlimited |
| **Voice quality** | Natural, Most pleasant | Natural | Robotic, Sometimes not appreciated by the user |
| **Multiple**  **featured voices** | Need high storage in that case | Need high storage in that case | Can produce multiple featured voices without any major changes |
| **Intelligibility** | High | Highly Intelligible | High |

*Table 1: Comparison between different categories of TTS*

References

[1] “Origins of Human Communication,” *MIT Press*.

[2] “FreeTTS 1.2.3 - A speech synthesizer written entirely in the Java(TM) programming language.” [Online]. Available: https://freetts.sourceforge.io/#what\_is\_freetts. [Accessed: 18-Jun-2020].