Indian Institute of Technology, Guwahati

Department of Computer Science and Engineering

Project Report on

Small Tutorial for Kids

Based on Speech Recognition System

Submitted to:

Prof. P. K. Das

Submitted by:

Prateekshya Priyadarshini (214101037)

Rohan Jaiswal (214101042)

For course fulfilment of CS566: Speech Processing

ACKNOWLEDGEMENT

This project is being submitted as a requirement for course fulfilment of CS566 - Speech Processing. It is a pleasure to acknowledge our sense of gratitude to Prof. P.K. Das who guided us throughout the project work. His timely guidance and suggestions were encouraging. We thank to the Teaching Assistants who were always helpful in clearing doubts. Finally, we thank to our classmates for the support.

1. Prateekshya Priyadarshini (214101037) 2. Rohan Jaiswal (214101042)

Contents

List of Figures

List of Tables

1   Abstract

This project is developed using C++/C. It can take a speech sample of a few seconds, preferably a single word, and display it's corresponding webpage which gives related information to the word. Initially it is developed for simple words which can be used as a tutorial for kids. But it can be expanded further for a bigger area of words. It uses the concepts of the famous Hidden Markov Model to store the properties of the speech sample and compare the new sample with these properties to detect which word has been spoken.

2   Introduction

2.1   What is Speech Recognition

Speech Recognition is a technique which is quite popular now-a-days. When we speak into a microphone which is connected to the computer/mobile, it converts it to a text file which contains some amplitude values. Those values are basically the deviation of the speech signal from X-axis. Then we can use this file, do some calculations which can detect which word has been spoken and then further steps can be taken as per the requirement. One such application is Alexa.

2.2   Our Project

This project uses a similar technique. There is a set of predefined words - aeroplane, ambulance, apple, autorickshaw, bicycle, bike, bus, car, cat, dog, scooter, tandem, train, tram, truck, orange and taxi. We can run the project and speak any of these words to see the related webpage. We can also train these words for new speakers. We can add new words which might take several minutes.

2.3   Future improvements

Since this project is developed using C/C++, it is difficult to run multiple things parallelly. Future improvements include development of this project using High Level Languages like Java or Python which can use multithreading concepts to run the model training part in background. This will remove the waiting time during training of new words.