Speech Processing Project Report

CSE(M.Tech) - IIT Guwahati

TITLE: - Interactive Quiz

Submitted By: -

Surbhi Jain(214101066)

Neha Afreen (214101034)

Submitted To: -

P.K. Das

(CSE Department)

Table of Content: -

Abstract

This document defines a set of evaluation criteria and test methods for speech recognition systems used in interactive session in which system will dictate the questions and options and user will record the answer. We used **Hidden Markov model (HMM**) for word recognition.

Introduction

The goal of this project is to define a set of evaluation criteria and test methods for the **interactive voice-based question answer-based session** and saying that spoken word is correct or not.

In this report we concentrate on the speech recognition programs that are human-computer interactive. When software evaluators observe humans testing such software programs, they gain valuable insights into technological problems and barriers that they may never witness otherwise. Testing speech recognition products for universal usability is an important step before considering the product to be a viable solution for its customers later. This document concerns Speech Recognition accuracy in contact searching and retrieving details, which is a critical factor in the development of hands-free human machine interactive devices. There are two separate issues that we want to test: word recognition accuracy and software friendliness. Major factors that impede recognition accuracy in the environment noise sources and system noise.

Speech Recognition: You speak into a microphone and the computer transforms the sound of your words into text to be used by your word processor or other applications available on your computer. The computer may repeat what you just said or it may give you a prompt for what you are expected to say next. This is the central promise of interactive speech recognition. You also had to correct any errors virtually as soon as they happened, which means that you had to concentrate so hard on the software that you often forgot what you were trying to say.

The new voice recognition systems are certainly much easier to use. You can speak at a normal pace without leaving distinct pauses between words. However, you cannot really use “natural speech” as claimed by the manufacturers. You must speak clearly, as you do when you speak to a Dictaphone or when you leave someone a telephone message. Remember, the computer is relying solely on your spoken words. It cannot interpret your tone or inflection, and it cannot interpret your gestures and facial expressions, which are part of everyday human communication. Some of the systems also look at whole phrases, not just the individual words you speak. They try to get information from the context of your speech, to help work out the correct interpretation.

Proposed Methodology

Basic requirements to develop this project are as follows:

✓ Windows OS

✓ Microsoft Visual Studio 2010

✓ C++ 11 integrated with VS2010

✓ Recording Module

With the availability of above software, we further procced in modelling the logic. The prerequisites of this project are

✓ Basic i/o operations on file

✓ Pre-processing of speech data

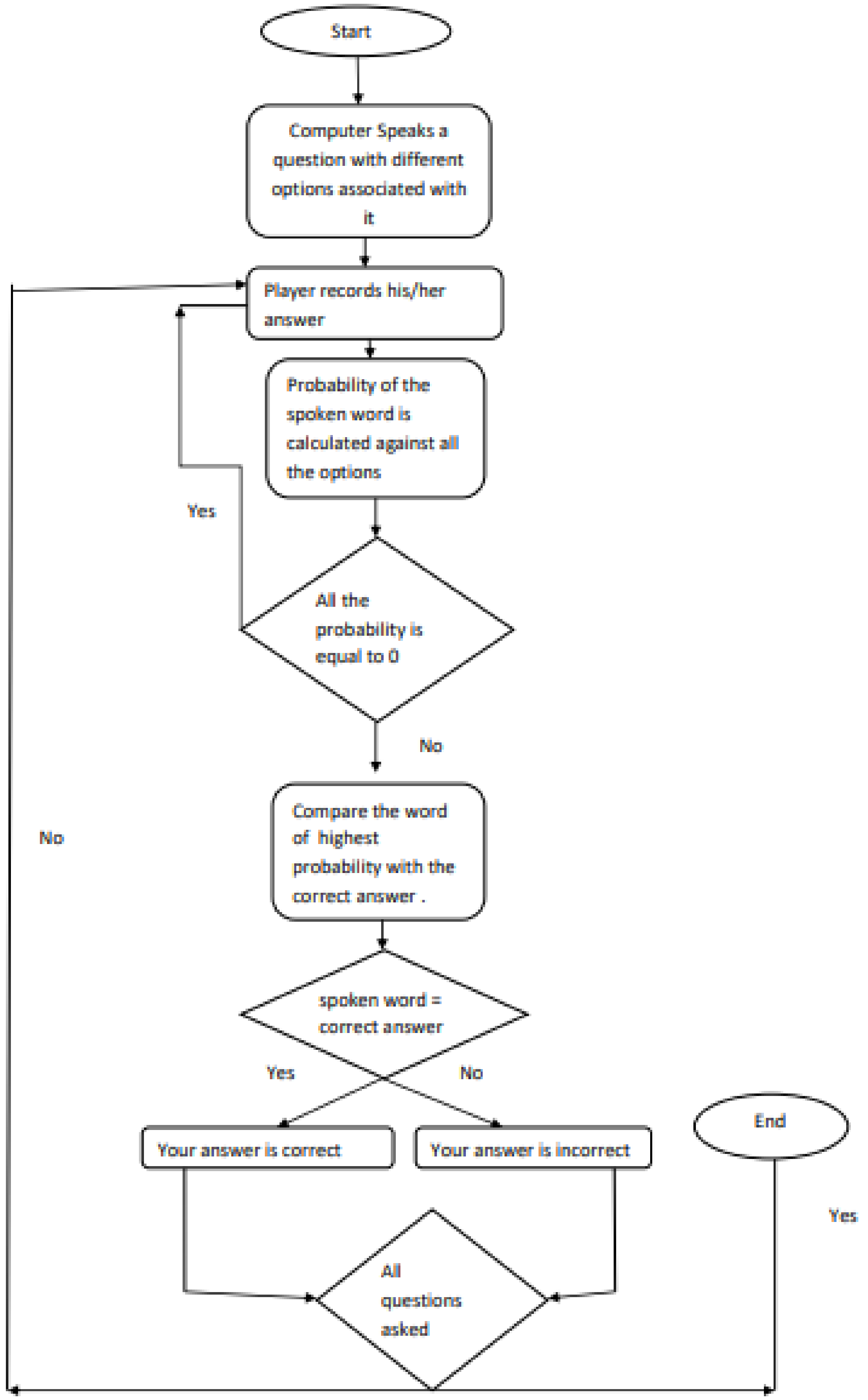
✓ Feature extraction

✓ Modelling of extracted feature

✓ Enhancing model

Above discussed topics are broadly elaborated in experimental setup section.

Flow Chart


Start
Computer Speaks a question with different options associated with it
Player records his/her answer
Probability of the spoken word is calculated against all the options
Yes
All the probability is equal to 0
No
No
Compare the word of highest probability with the correct answer .
spoken word = correct answer
Yes
No
Your answer is correct
Your answer is incorrect
End
Yes
All questions asked

EXPERIMENTAL SETUP

This project is divided into following modules:

1. Training Module
2. Testing Module
3. Live Training Module

1. Training Module:

The flow for training over data is as follows:

a. Record 20 utterances for each word which are used in project.

b. Extract frames of speech part from these recordings with 2-3 frames silence at beginning and end of word

c. Use front end analysis to get observation sequence from the speech.

d. Pass this observation sequence to HMM for model designing by using Bakis model(feed forward model) as initial model.

e. Now enhance the model using HMM re-estimation algorithm (Baum welch).

Now reference models for each word are ready for our project. The training of data is not integrated with GUI application. This is different module which will just evaluate reference model

2. Testing Module

We need search the word by speaking into the microphone.

The flow of testing is as follows:

a. Live recording of data is done on clicking on record answer button .

b. Testing will be done on the data with pretrained models.

c. Verifying the answer with 4 options using in that question.

d. If verification is successful then show that your answer is correct.

e. If verification fails then show that your answer is correct.

3. Live Training Module

The flow for live training over data is as follows:

a. Put the word and directory from the list mention in that page .

b. Click on train button.

c. Extract frames of pre recorded words.

d. Using front end analysis to find the observation sequence.

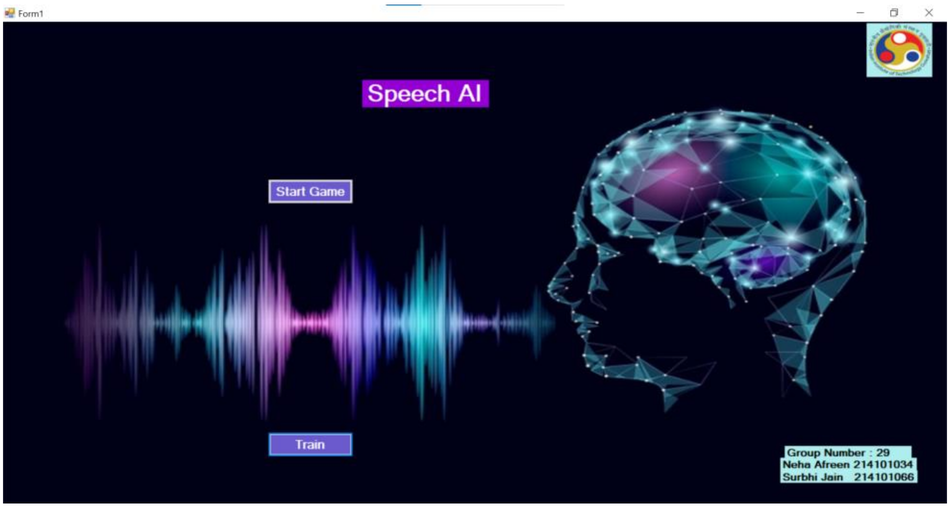
e. Pass this observation sequence to HMM for model designing by using Bakis model(feed forward model) as initial model.

f. Now enhance the model using HMM re-estimation algorithm (Baum welch).

Now the reference models for each word are ready and saved in directory.

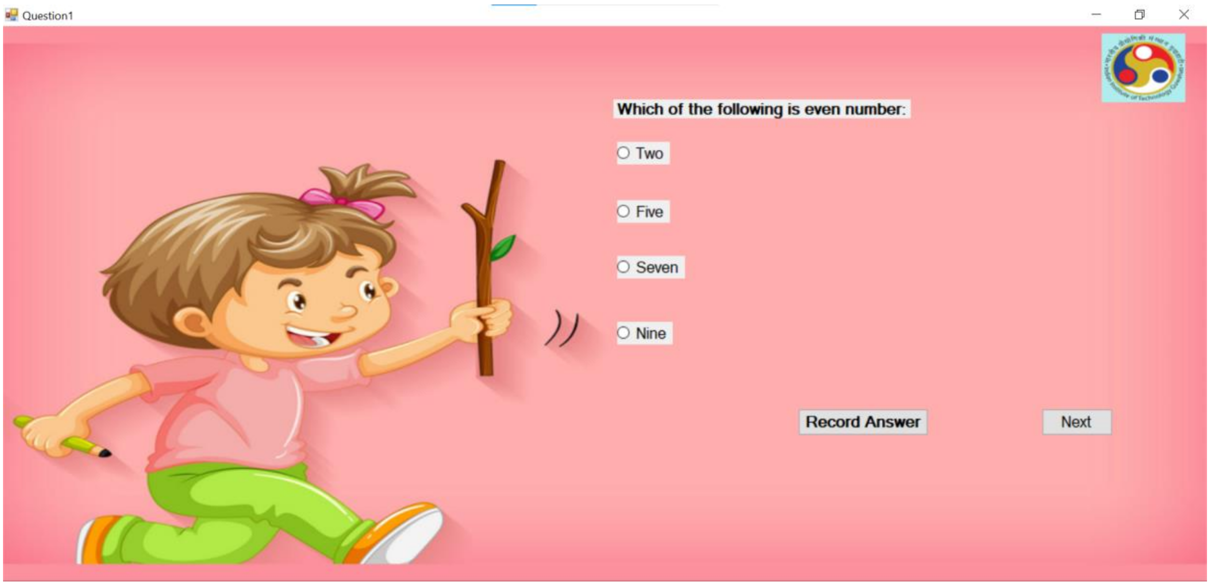
USER INTERFACE

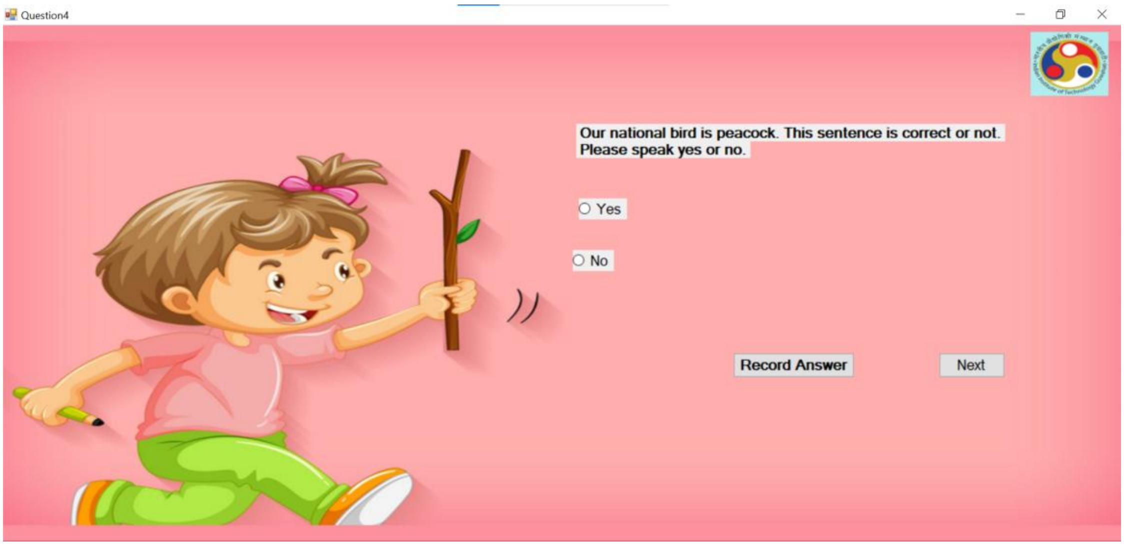
Front Screen:

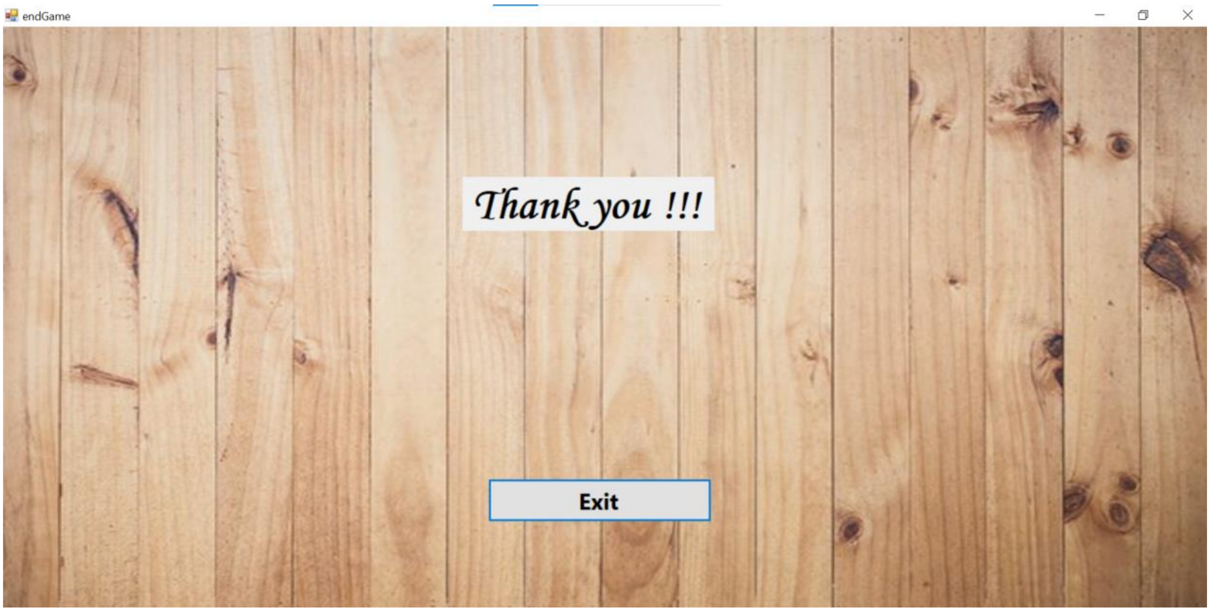


Start Game:

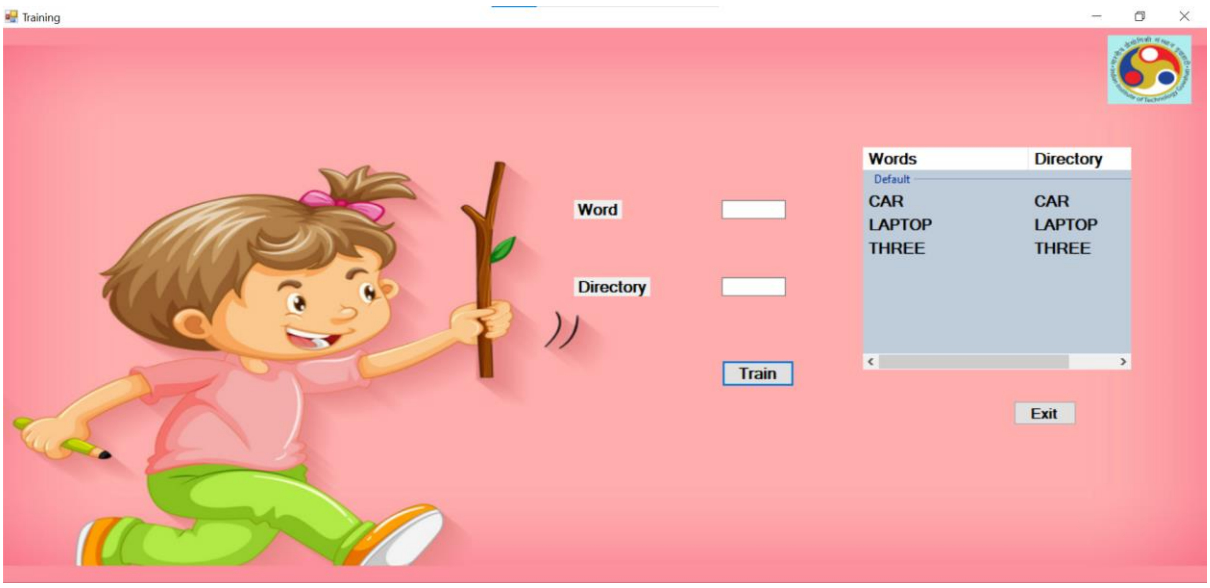
We have 4 objective type question in this game.




Question4
Our national bird is peacock. This sentence is correct or not.
Please speak yes or no.
Yes
No
Record Answer
Next


endGame
Thank you !!!
Exit

Train:


Training
Word
Directory
Train
Words
Directory
Default
CAR
CAR
LAPTOP
LAPTOP
THREE
THREE
Exit

Result

We have 4 objective type questions in this game and for each correct answer we are checking with reference models of the options which are already stored in the files.

Based on the probabilities from the models of the question option , it will tell that answer is correct or not.

System will speak and show whether answer is correct or not.