

Indian Institute of Technology, Guwahati



Department of Computer Science and Engineering

Project report

On

“Speech based Captcha Solver”

Based on

Speech recognition system

Course: CS566 Speech Processing

Submitted to

Prof. P. K. Das

Submitted by:

Gaurav Kumar (214101018)

Sanjeev Kumar (214101048)

TABLE OF CONTENT

1. Abstract
2. Introduction
3. Proposed Methodology
4. Experimental Setup
5. Project Snapshots
6. Result

ABSTRACT

This document defines a set of evaluation criteria and test methods for speech recognition systems used in solving captcha based on speech. This report is on the project Speech Based Captcha Solver. We have used HMM based Voice recognition technique to recognize the speech with a high accuracy. The project works well with 85 percent accuracy on live data and 98 percent on offline data. We have used speech based recognition system in logging in any application and restricting bot access.

INTRODUCTION

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.

But what is speech recognition?

Speech recognition works like this. 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.

The goal of this project is to define a set of evaluation criteria and test methods for the interactive voice recognition systems for searching contact and retrieving corresponding details for successful search.

PROPOSED METHODOLOGY

Basic requirements to develop this project are as follows:

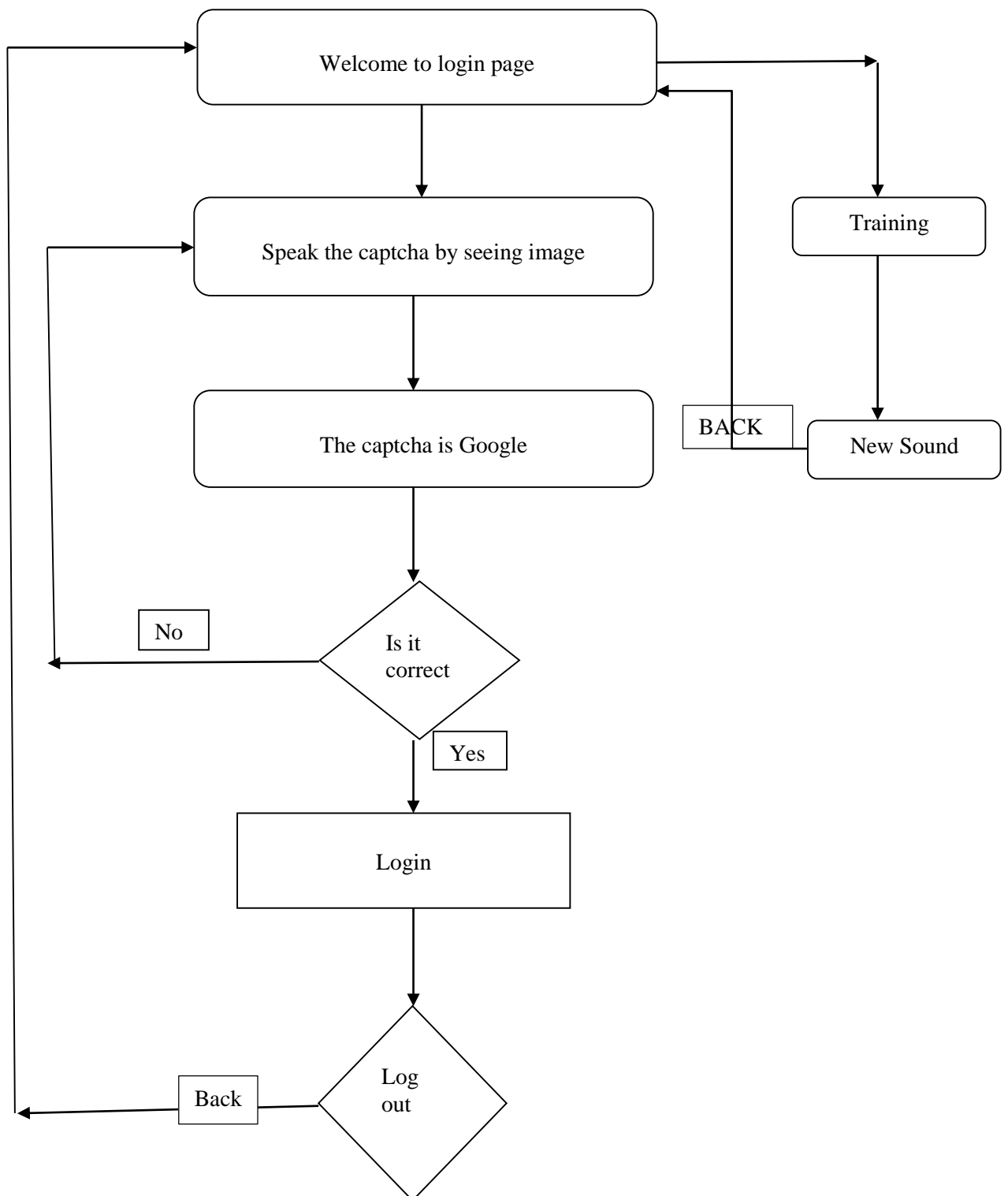
- ✓ Windows OS
- ✓ Microsoft Visual Studio 2010
- ✓ C++ 11 integrated with VS2010
- ✓ Recording Module

With the availability of above software, we further proceed 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.

With the availability of above tools, we further proceeded. Below is the flow chart for our project



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:

- i. Record the data as 20 utterances of each captcha word
- ii. Extract frames for every utterance
- iii. Using local distance analysis (in vector quantization) calculate the observation sequence.
- iv. Pass this observation sequence to HMM for model designing.
- v. Now enhance the model using HMM re-estimation algorithm.

Now reference model is 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

System will give instruction what is going on and user is required to follow it.

The flow of testing is as follows:

- i. Live recording of data is done when system instruct.
- ii. Testing the data with pretrained models.
- iii. Verifying the captcha detected with user.
- iv. If verification is successful login will happen.
- v. If verification fails, record the input again.

Project Snapshots

1. Entry Page (Log In Page)

Form1

Enter Email Id.

Enter Roll Number.

zero

Press record and speak...

☐ I am not a Robot

Training

Login

Restart

2. For unsuccessful captcha attempt

Form1

Enter Email Id.

Enter Roll Number.

OLA

Recognised: one

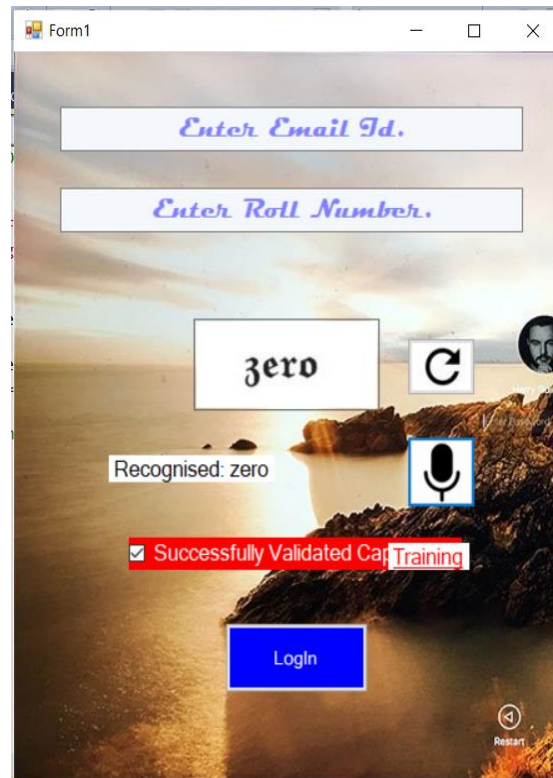
☒ Try Again....

Training

Login

Restart

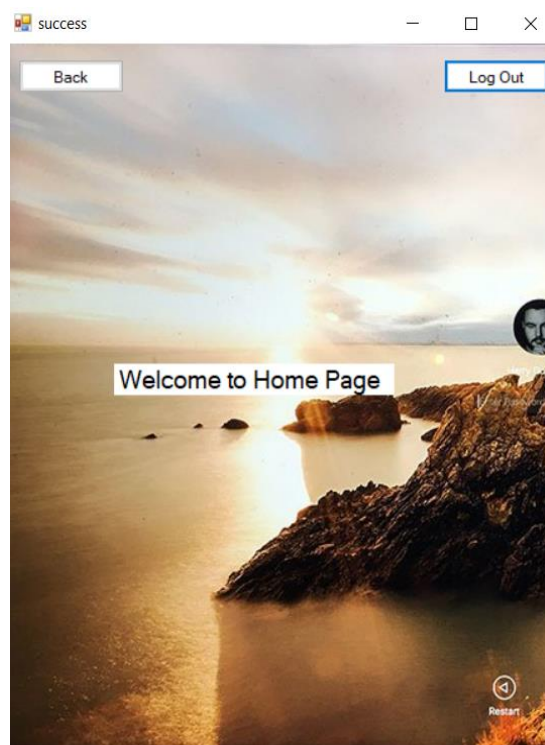
3. For Successful captcha recognition



The screenshot shows a login window titled "Form1" with a background image of a sunset over a rocky coastline. The form contains the following elements:

- Input field: *Enter Email Id.*
- Input field: *Enter Roll Number.*
- Captcha display: A box showing the word "zero" with a refresh icon to its right.
- Recognition status: A label "Recognised: zero" and a microphone icon.
- Validation message: A red banner with a checkmark icon and the text "Successfully Validated Cap Training".
- Login button: A blue button labeled "Login".
- Restart button: A small circular icon with a left arrow and the text "Restart" at the bottom right.

4. For Successful captcha recognition



The screenshot shows a success window titled "success" with the same sunset background. The page contains the following elements:

- Navigation buttons: "Back" and "Log Out" at the top.
- Welcome message: A label "Welcome to Home Page".
- Restart button: A small circular icon with a left arrow and the text "Restart" at the bottom right.

5. For Training on new data

training

Back

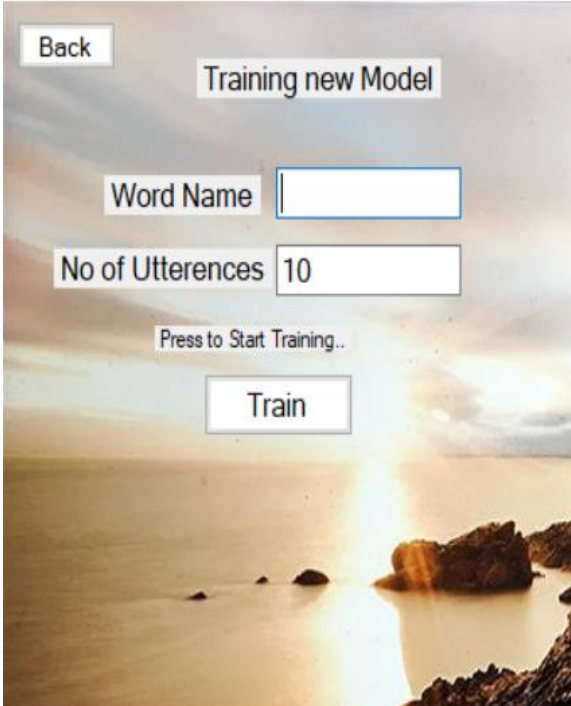
Training new Model

Word Name

No of Utterences

Press to Start Training..

Train



RESULT

The complete login system based on voice based captcha is Implemented. We have started with 15 initial models to identify the captcha. New model can be trained based on new sound based on training module.

The accuracy of Voice based captcha solver is 98 percent on offline data and 85 percent on live data. Thus concluding that HMM based speech system works pretty well and faster.