

Indian Institute of Technology, Guwahati



Department of Computer Science and Engineering

Project report

On

“Speechify ReCAPTCHA App”

Based on

Speech recognition system

Course: CS566 Speech Processing

Submitted to

Prof. P. K. Das

Submitted by:

Aishwarya Gupta(214101019)

Shambhavi Shanker(214101051)

TABLE OF CONTENT

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

ABSTRACT

This document defines various processing and test methods required for the development of speech recognition based applications like speech based ReCAPTCHA app. This report is on the project which takes the ReCAPTCHA's input from the speaker(user) in the form of speech.

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 login system using speech based captcha.

PROPOSED METHODOLOGY

Basic requirements to develop this project are as follows:

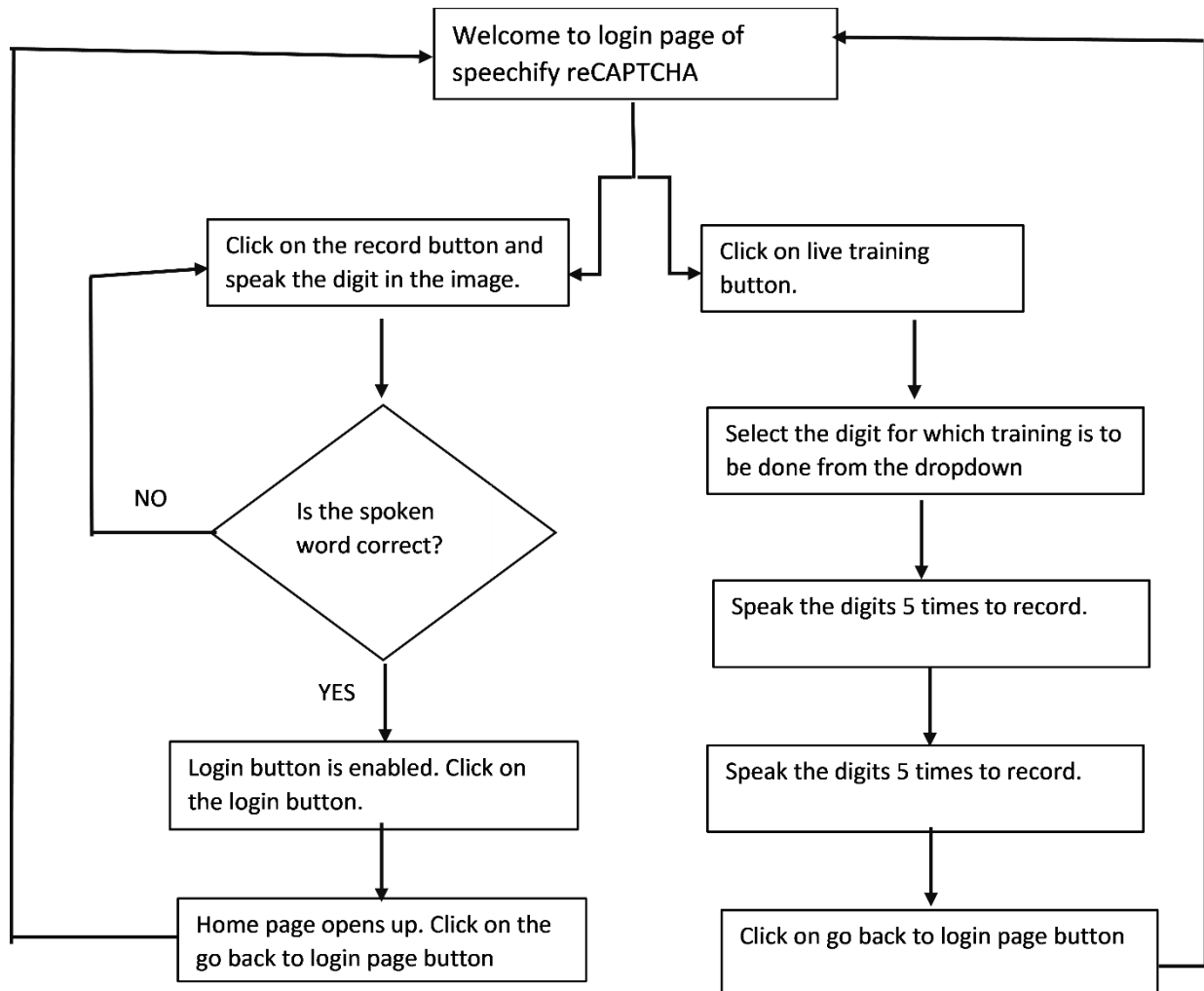
- ✓ Windows OS
- ✓ Microsoft Visual Studio 2010
- ✓ Recording Module

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 digit (0-9).
- ii. Normalize the data and then extract frames for each utterance.
- iii. Compute the observation sequence by vector quantization technique.
- iv. Pass this observation sequence to HMM for model training.
- v. Now enhance the model using HMM re-estimation algorithm.

Now the final model for each digit 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 instructs.
- ii. Testing the data with pretrained models.
- iii. Verifying the captcha with the word spoken by the user.
- iv. If spoken word is same as the word in the image, the user is allowed to login.
- v. If verification fails, user can try again and has 3 attempts in total.

3. Live training module

- i. The user is asked to select the word for which recording is to be done.
- ii. Then the user has to speak that word 5 times.
- iii. The word is recorded and stored in a folder.
- iv. A new model is trained for that user for that particular word.

RESULT

A speech based captcha solver has been implemented in this project.