Indian Institute of Technology, Guwahati

Department of Computer Science and Engineering

Project report

On

Speech based Captcha Solver

(based on speech recoginition system)

Course: CS566 Speech Processing

Submitted to:

Prof. Pradip Kumar Das

Submitted by: -

Samiul Haque (224101044)

Vishal Chinchkhede (224101057)

Amit Kumar (224101006)

Index

ABSTRACT

The assessment criteria and testing procedures for voice recognition systems used in captcha solving are described in this documentation. This report focuses on a project that recognizes user-spoken words in captcha text.

INTRODUCTION

In this research, we focus on voice recognition program that relies on human-computer interaction. Software testers may learn a lot about technological issues and impediments when they see people using these software applications in real-world settings. Testing speech-based products under different conditions is a required and obligatory step. Speech-based products provide significant client problems if they are not tested. Based on a speech-based automated captcha solution, this documentation enables hand-free human-computer interaction. Two key concerns must be maintained while the project is being developed: accuracy and product usability.

The method through which a computer (or other sort of machine) recognizes spoken words is called speech recognition. In essence, it is speaking to your computer AND having it understand what you are saying. The fundamental definitions required to comprehend voice recognition technology are as follows –

1. Utterance

2. Dependence on Speaker

3. Vocabularies

4. Accuracy

5. Training

Using the latest speech recognition technologies is undoubtedly more simpler. You don't need to hesitate between sentences; you may speak normally. However, despite what the producers suggest, "natural speech" cannot genuinely be used. You must speak clearly, just like you would when talking to a dictaphone or leaving a voicemail for someone. Keep in mind that the computer relies only on your spoken words. Your movements and facial expressions, which are a crucial aspect of normal human communication, are all incomprehensible to it. It also is unable to understand your tone or inflection. Some of the systems analyze more than just the words you speak; they can analyze whole sentences. In order to determine the proper interpretation, they attempt to glean information from the context of your speech. The goal of this project is to define a set of evaluation criteria and test methods for the interactive voice recognition systems for capture solving used in hand free human computer interaction.

This project's objective is to provide a set of assessment standards and testing procedures for interactive speech recognition systems for capture-based human-computer interaction.

FLOWCHART
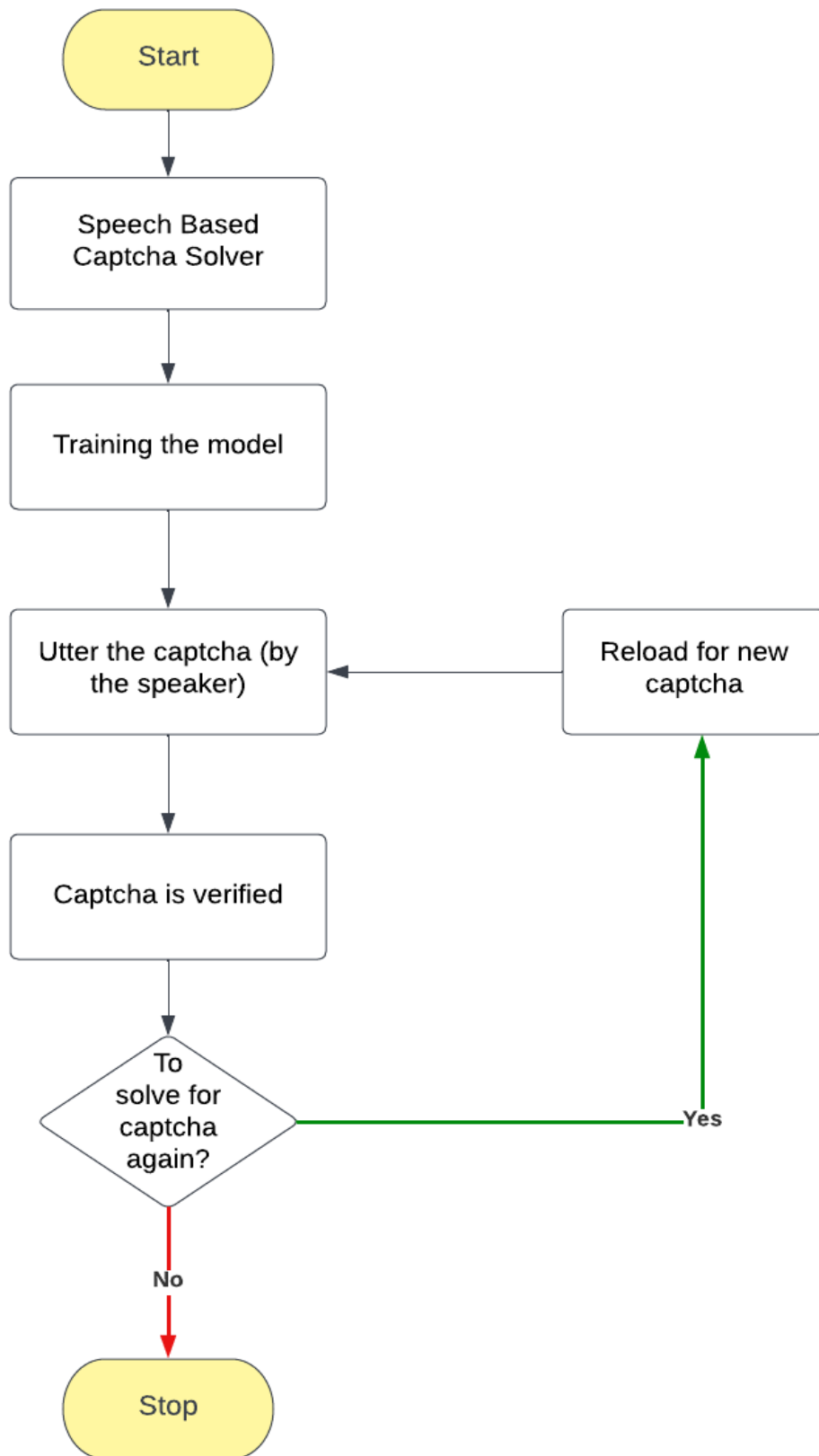
SUGGESTIVE METHODOLOGY

The following are the basic criteria to build this project:

- Windows OS

- C++ 11 integrated with VS2010

- Microsoft Visual Studio 2010

- Recording Module

With the aforesaid tools readily available, we continued with the logic modelling. The conditions necessary for this project are:

- Basic input/output operations on file
- Processing speech data beforehand(pre-processing)
- Feature extraction
- Extracted feature modelling
- Improving the model

In the section on the experimental setup, the previously mentioned themes are substantially expanded

EXPERIMENTAL SETUP

This project is divided into following modules:

- Training Module
- Testing Module (Seen Data)
- Testing Module (Unseen Data)
- Live Testing
- Live Training Module (Developer's Option)

1) Training Module:

Following is the flow for training on data:

a) Keep track of the data as 20 repetitions of each word.

b) To each utterance, extract frames

c) Calculate the observation sequence using local distance analysis (in the context of vector quantization).

d) Send this observation sequence to HMM so that it can design a model.

e) Utilizing the HMM re-estimation algorithm, improve the model now.

Our project's reference model is now ready. There is no integration between the training of data and GUI applications. This is a distinct module that will only assess the reference model.

2) Testing Module:

User must follow instructions provided by the system to understand what is happening. The testing process is as follows:

a) Data is recorded in real time as the system directs.

b) Using trained models to test the data.

c) Reload for a fresh capture.

RESULT

The captcha that are already kept in a data file are what we are obtaining. The need to fetch data based on speakers is carried out satisfactorily.

Accuracy:

1. On seen data (data used to train the model): - 100%
2. On unseen data: - 97%
3. During live testing, at least 75% (works only with the voice of a single speaker on whose voice the training was done)

ADDITIONAL ROOM FOR IMPROVEMENT

It is intended for this live captcha solver module to be a component of a bigger system. Therefore, in the future, we hope to include this module into a bigger system, such as a login system or another system that employs a captcha.