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: -

Rittick Mondal (214101041)

Anamitra Mukherjee (214101007)

TABLE OF CONTENT

ABSTRACT

This document defines a set of evaluation criteria and test methods for speech recognition systems used in capture solving. This report is on the project which detect the word (captcha text) spoken by user.

INTRODUCTION

In this report we are concentrate on recognition of speech programs, which is based on human computer interaction. When software evaluators observe humans testing such software programs, they gain valuable insights into technological problems and barriers otherwise may that never be explored. Testing of speech based product in various condition is mandatory and necessary step .Without testing speech based products create very much problem for customer. This documentation is based on speech based automatic capture solver, which helps for hand free human computer interaction. During development of the project we have to maintain two main issue, one is accuracy and another is user friendliness of product.

Speech recognition is the process by which a computer (or other type of machine) identifies spoken words. Basically, it means talking to your computer, AND having it correctly recognize what you are saying. The following definitions are the basics needed for understanding speech recognition technology –

1. Utterance

2. Speaker Dependence

3. Vocabularies

4. Accuracy

5. Training

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 capture solving used in hand free human computer interaction.

PROPOSED METHODOLOGY

Basic requirements to develop this project are as follows:

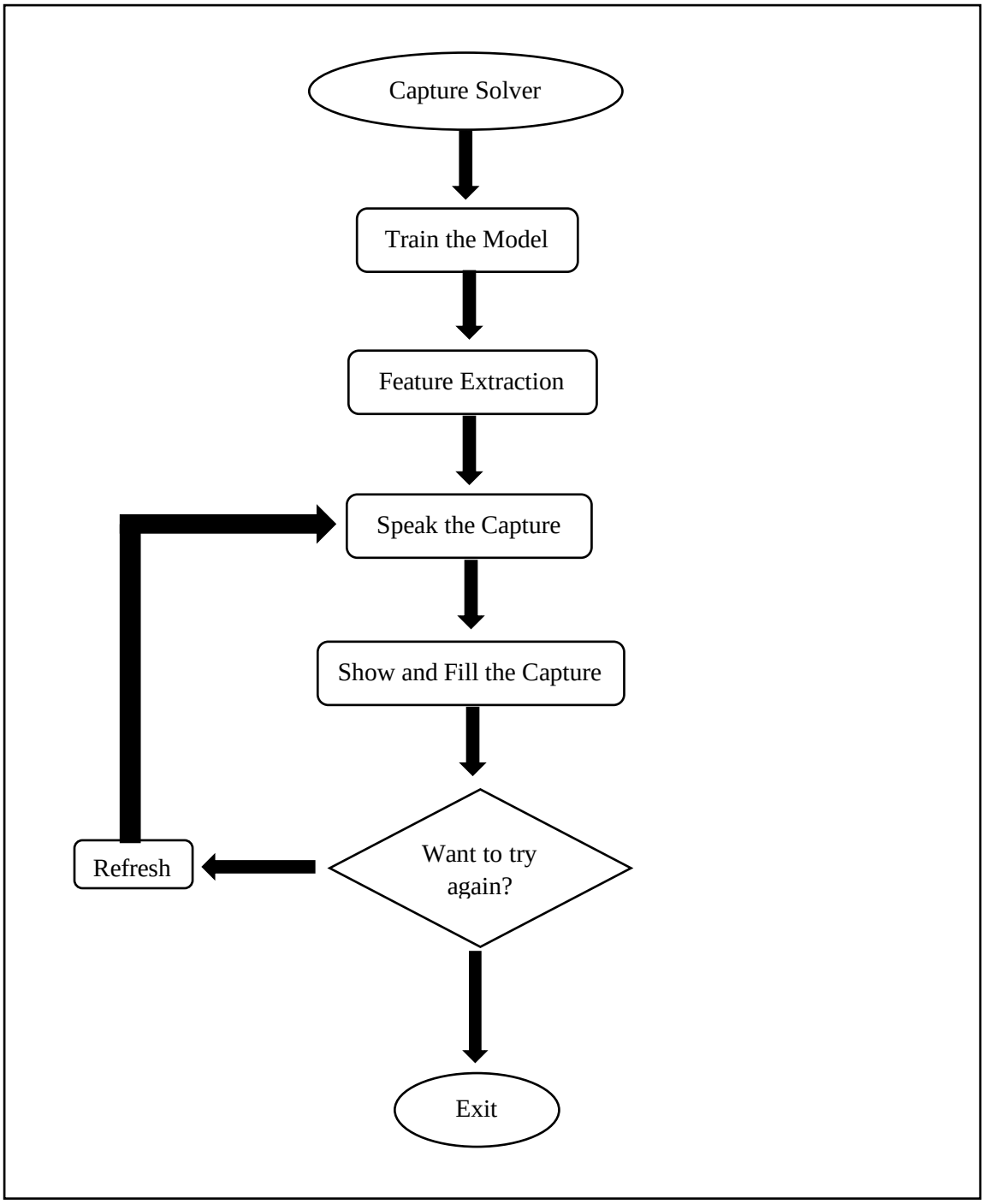
- Windows OS
- Microsoft Visual Studio 2010
- C++ 11 integrated with VS2010
- Recording Module

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:-


Capture Solver
Train the Model
Feature Extraction
Speak the Capture
Show and Fill the Capture
Want to try again?
Refresh
Exit

EXPERIMENTAL SETUP

This project is divided into following modules:

1. Training Module
2. Testing Module (Seen Data)
3. Testing Module (Unseen Data)
4. Live Testing
5. Live Training Module (Developer's Option)

1. Training Module
   The flow for training over data is as follows:
   i. Record the data as 20 utterance of each 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 pre-trained models.
   iii. Refresh for new Capture

RESULT

We are getting the captures that are already stored in a data file. Fetching of data based on speakers is requirement is successfully implemented.

Accuracy:-

1. On seen data (data on which model was trained): - 100%
2. On seen data: - 97%
3. Live Testing: - Greater than 75% (works only with the voice of a single speaker on whose voice the training was done)

FURTHER SCOPE FOR IMPROVEMENT

This live captcha solver module is meant to be a part of a larger system. So, in future we would like to integrate this module to a larger system like a login system or any other system that uses a captcha.