Indian Institute of Technology, Guwahati

Department of Computer Science and Engineering

Project report

On

“Speech based Contact Search App”

Based on

Speech recognition system

Course: CS566 Speech Processing

Submitted to

Prof. P. K. Das

Submitted by:

Ira Bhist (194101019)

Prachi Srivastava (194101037)

Shashank Maurya (194101044)

Sudeshna Chaudhari (194101050)

TABLE OF CONTENT

ABSTRACT

This document defines a set of evaluation criteria and test methods for speech recognition systems used in searching and retrieving contact details. This report is on the project which detects the contact name and show its details

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:

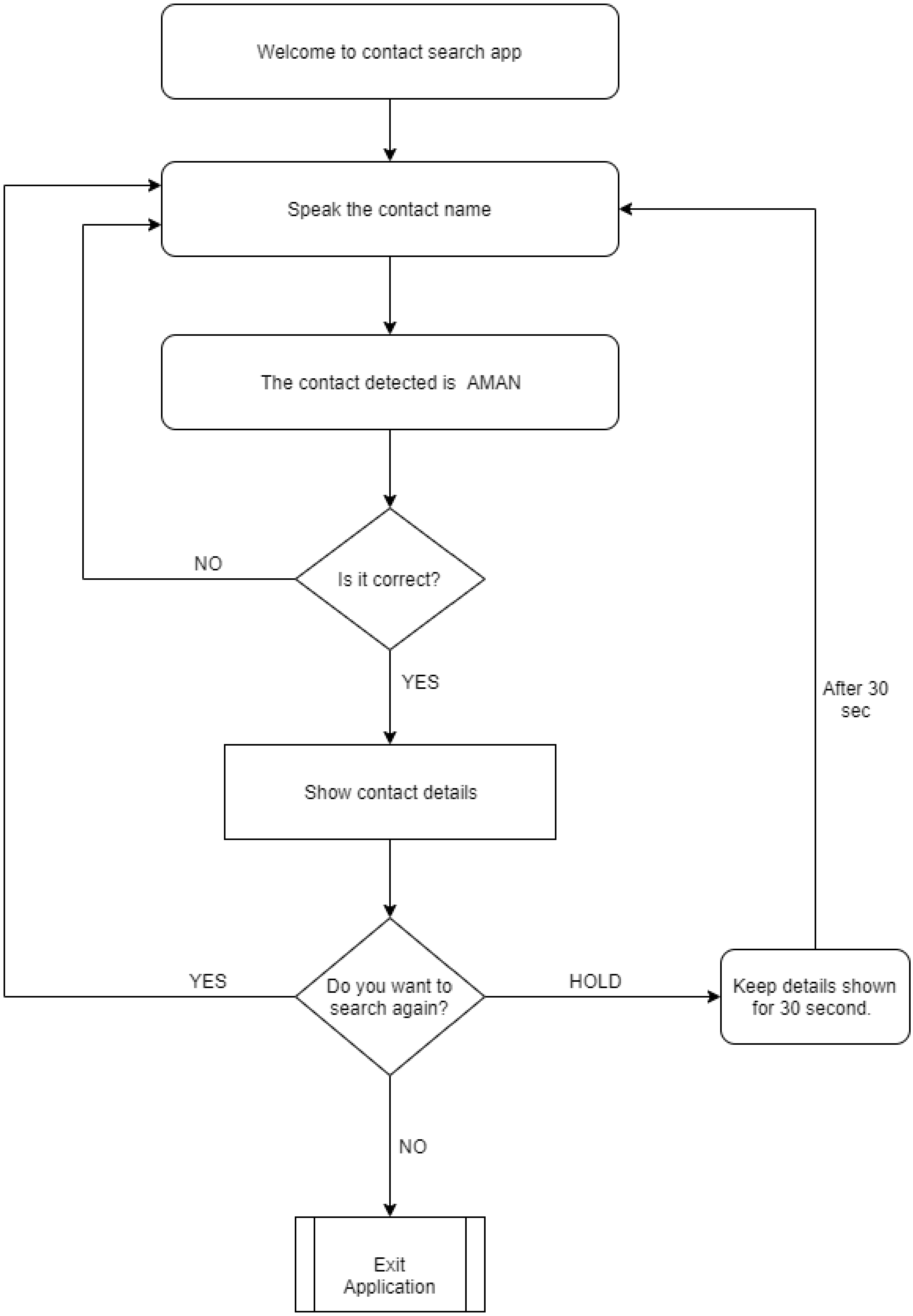
- ✓ Windows OS
- ✓ Microsoft Visual Studio 2010
- ✓ C++ 11 integrated with VS2010
- ✓ Recording Module

With the availability of above software, we further 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.

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 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 pretrained models.
   iii. Verifying the contact name detected with user.
   iv. If verification is successful display contact details.
   v. If verification fails, record the input again.

RESULT

We are getting contact details of 10 contacts which are already stored in a data file. Fetching of data based on speakers is requirement is successfully implemented.