# Indian Institute of Technology, Guwahati



Department of Computer Science and Engineering Project report

On

# “Voice Controlled Alarm Clock”

Based on

### Speech recognition system

Course: CS566 Speech Processing

Submitted to Prof. P. K. Das

Submitted by:

Kajal Khobragade (224101028)

**TABLE OF CONTENT**

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

## ABSTRACT

#### This document defines a set of evaluation criteria and test methods for speech recognition systems used in setting, stopping and snoozing a voice controlled alarm clock. This report is on the project which detects the voice command given by the user and sets an alarm for that particular time.

**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 setting alarm in the system almost hands free through using your voice only, 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 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 successfully setting, stopping and snoozing an alarm only through using your voice commands.

## PROPOSED METHODOLOGY

Basic requirements to develop this project are as follows:

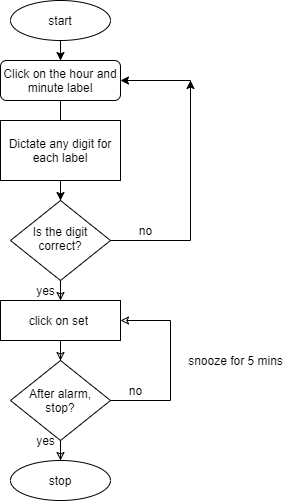
* Windows OS
* Microsoft Visual Studio 2010
* C++ 11 integrated with VS2010
* Recording Module

#### 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
4. Training Module

The flow for training over data is as follows:

* 1. Record the data as 20 utterance of each digit between 0-9 as well as recordings of basic operations STOP, SNOOZE, AM and PM.
  2. Extract frames for every utterance
  3. Using local distance analysis (in vector quantization) calculate the observation sequence.
  4. Pass this observation sequence to HMM for model designing.
  5. Now enhance the model using HMM re-estimation algorithm.

Now reference model for each digit/word 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 and can be used to train for the new user.

1. Testing Module

System will give instruction what is going on and user is required to follow it. The flow of testing is as follows:

* 1. Live recording of data is done when system instruct.
  2. Testing the data with already existing models.
  3. Verifying the time dictated by the user.
  4. If verification is successful set an alarm for that particular time in the system.
  5. If verification fails, record the input again.

## RESULT

We are able to successfully set, stop, and snooze an alarm for whatever time the user desires to dictate. To keep the experiment which is more inclined to practical usage, we have kept the alarm timings at 10 minute increments as most people usually set there alarms in 10 minutes intervals and to keep the models as simple and recognition can be done more precisely.

Example: - 11:30pm, 02:10am, 05:00pm etc.

That 30 minute increment detection is done in order to decrease the time taken as we don’t have to train and compare with 60 models. The “seconds” label is always kept at 00 which can’t be manipulated by any command.

Uttering “Snooze” will increase the alarm timer by 5 minutes which is kept as default.

Uttering “Stop” will stop the alarm.