International Journal of Science and Engineering Applications
Volume 8–Issue 07,221-225, 2019, ISSN:-2319–7560

# Algorithmic Design Analysis of Voice Recognition System

Kamalu U. A
University of Portharcourt
Rivers State, Nigeria

Dike Precious
University of Portharcourt
Rivers State, Nigeria

**Abstract**: This paper is on Algorithmic design analysis of voice recognition systems, it involves the design and analysis of voice recognition system using algorithms. A modular design approach was adopted, the designed submodules are incorporated together to form a complete unit, the designed system is general and can be adopted to fit into various voice recognition systems. It can also be programmed using any suitable programming language like C++, PHP etc. The algorithm accepts voice as input analyzes the voice of interest, retrieves the parameters from the voice, performs a database input matching/search for a user who has similar records and then displays an output of the result.

**Keywords**: Voice Recognition, Speech Recognition, Algorithm, Machine, Artificial Intelligence

## 1. INTRODUCTION

An Algorithm is an unambiguous specification of how to solve a problem or class of problems. Algorithms can perform calculation, data processing, automated reasoning, and other tasks. Starting from an initial state and initial input which could be empty, the instructions describe a computation that, when executed, proceeds through a finite number of well-defined successive states, eventually producing an "output" and terminating at a final ending state. Design analysis is the study of the characteristics (form, structure, construction, features, etc.) of the design of a system using a systematic process to find out certain other characteristics of the thing, in simple terms it is using the known to find the unknown, it focuses on studying and reviewing the details of the design in order to answer questions about the system such as will it break, when will it break, how long will it last etc. Voice Recognition is the identification of a person from the characteristics of the voice.The term voice recognition can refer to speaker recognition, recognizing the speaker can simplify the task of translating speech in systems that have been trained on specific voices or it can be used to authenticate or verify the identity of a speaker as part of a security process. Speaker recognition has a history dating back some four decades and uses the acoustic features of speech that have been found to differ between individuals. A system which performs voice recognition is referred to as a Voice Recognition System.

This paper will cover the step by step process involved in the identification of a person from the characteristics(frequency, Amplitude) of the voice in a voice recognition system.

## 2. METHODOLOGY

The modules involved in the design are the input module which is the voice or text, the database search or input matching module, parameter retrieval module,voice analysis module based on frequency and amplitude parameters and the display of the result.

Based on the above the system is split into workable smaller units:

- Input Module
- Analysis Module
- Parameters Modules
- Database Module
- Output Module
- General Modules

All the functionalities of the modules will be setup as submodules that are either public or private to other modules so as to enable interconnectivity among modules or resource sharing.

**Input Module:**

The system's input will be a microphone placed at a position where the user can speak directly to it and the microphone has to be able to cancel noise in case the user is in a noisy area and a keypad for a user to type meaning the system input might need to be text or voice and either the keypad or microphone is used depending on the needed input.

This module simply accepts input from either the microphone or the keypad and should be able to access all public sub modules from other modules. Its submodules are

- Mic Input sub module
- Text Input sub module

the Mic Input Sub module which can retrieve the voice from the mic and it is a private submodule meaning only the Input module has access to it while the Text input sub module accepts text input from the user and it is also a private sub module.

**Analysis Module:**

This module handles filtering and recording of voice and the various methods of analysis of the recorded voice which are sampling and frequency domain analysis hence its sub modules are

- Filter Sub module
- Record Sub module
- Sampling analysis Sub module
- Frequency domain analysis Sub module

The ecological importance of human voice processing for listeners is obvious and it is important to note that voice does not only convey speech, it also provides information on user's gender, identity, and emotional state [1]. The Recording submodule has to be a public sub module because it will be used at some point by the Input Module while the Filter, Sampling and Frequency domain analysis sub modules are private sub modules.

The Filter sub module is for removing of unwanted sounds/noise in the voice data gotten form the Input module and neuroimaging studies confirm that there is neural circuitry just for vocal sounds [2]. Also, the system should automatically separate voice sound and non-voice sounds which would be useful for further filtering of noise from the voice and speech-recognition or keyword-spotting could ignore the sections of a noisy part that are dominated by non-vocal sounds, these are all the features that the Filter sub module has.


Fig 1. Example of a Sampled Voice signal

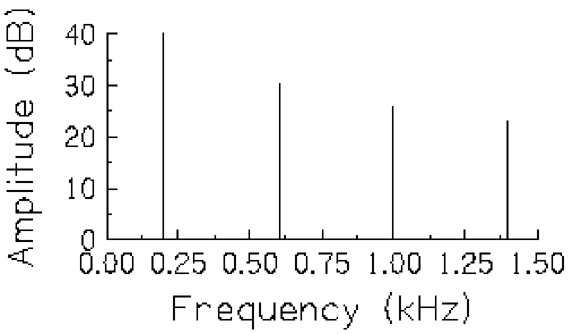


Fig 2. Example of a Frequency Domain signal

Parameters Module:

This module handles the retrieving of needed parameters from the Analysis Module and each parameter to be retrieved is done by a dedicated submodule and they are

Frequency Sub module

Amplitude Sub module

Note that all the submodules here are private because they are only needed by the Parameters Module and based on their name it is easy to know which parameter the sub module will be retrieving, the Frequency sub module will be retrieving the frequency of the Voice from the Frequency Domain data, Amplitude sub module will be retrieving the amplitude from the Sampled data.

Database Module:

This modules handles database related processes and such as saving of data and retrieving of data and the system will need just 2 sub modules:

Save Submodule

Get Sub module

The Save submodules INSERT's records in to the database while the Get submodule performs the GET FROM command on the database for getting similar records which is useful for our search and it is important to note that both submodules are private.

Table - users

Fields - id,full_name,sex,created_at,updated_at

Table - users_voice

Fields - id, user_id, voice_file, tone, frequency, amplitude, created_at, updated_at

Above is a visual illustration of the tables and fields of our basic database, the full_name and sex of the users table is entered during enrollment/training while created_at and updated_at are date-time stamps of when the record was inserted and last updated respectively. The user_id of the users_voice table is the key that relates the user record to his voice record on the users_voice table meaning the system has to use a relational database with each table having a primary key and id field being an integer and auto-increment. The voice_file is a blob field that holds the audio voice record, the frequency field holds the frequency of the voice, the amplitude field holds the amplitude of the voice and created_at and updated_at are also date-time stamps of when the record was inserted and last updated respectively.

Output Module

This module handles communicating results back to the user and it has 2 sub modules namely

Buzzer Sub module

Display Sub module

Where the buzzer sub module makes a buzzing sound to get the users attention the Display sub module displays the response/result on a screen and they are both private sub modules.

General Module

This modules is for handling operations that could not be part of the Above Modules and they are equally as important as the above modules as they perform operations that makes things easy for the user, the sub modules are

Sentence generator

Error handling

The sentence generator submodule handles generating of sentences used during the training process for the user to read out and it is limited to 3 sentences while the error handling sub module is to handle any unforeseen errors in the system by generating a system log that can be accessed by the administrator and it gives details of the what was going on during which the error occurred such as

The current module

The current sub module

The time

The process eg training/identification

The details of the last input or output data it has

with the above details it give the administrator all the information to aid in the debugging and helps make the identification of the error faster. All submodules in the General module are public hence they are available to all modules.

With this breakdown of the entire process into modules and submodules the system can be easily understood and the functionality of the system gets clearer. It is important to note that there are some functionalities which more than one module might need at a particular time and an example is voice recording hence the Recording submodule is be accessed by more than one module which is the reason why it is a public sub module and for clarity sake all actions of each module are setup as sub modules.

It is important to note also that the system needs to have a way of collecting users data and keeping it so users can be identified later on, this means the system will require "training" also called "enrollment" by each user where an individual speaker reads text or isolated vocabulary into the system. The system analyzes the reader's specific voice and uses it to fine-tune the recognition of that user's speech, which will result in increased accuracy. This is compulsory because the system being developed is a "speaker dependent" system and during the process the reader has to read the text in 3 different tones namely

Low Tone

Mid Tone

High Tone

This is to ensure the system can detect a user when the user speaks on a different tone because a users mood can affect the tone of the users voice and with this the system can easily detect the user even when his tone changes either due to mood swings, emotional troubles or vocal problems.Tone will vary from user to user depending on how a user reacts to an emotion, some studies show an increase in intensity,increase in fundamental frequency [3].

The flow of the system during training or enrollment is different form the flow during user identification due to the fact that during a training the database module should run the Save submodule while during identification it should run a Get Sub module, below are the system flow for the various process.

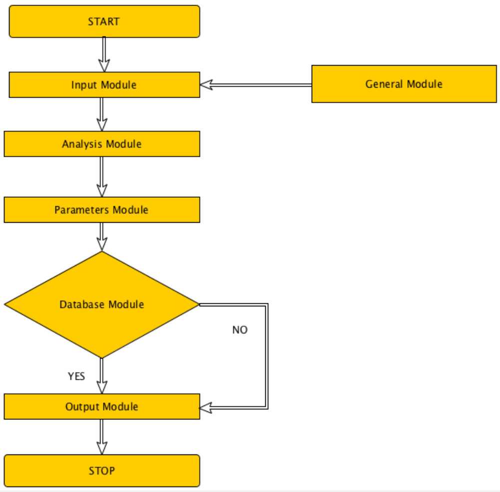


Fig 3. Flow chart representation of each module

IDENTIFICATION PROCESS:

The Input Module

The system receives the voice of the user from the microphone using the MIC Input sub module and sends the voice data to the Analysis Module

Analysis Module

The system at this point filters and analyzes the voice from the input module in order to make certain parameters or variables available and the voice is first recorded as an audio file using the RECORD sub module and then passed to our submodules to analyse the audio in different categories so the system can extract required variables in the next stage.

Parameters Module

The system retrieve the important parameters from the analysis conducted in the Analysis module and the following variables will be retrieved

Frequency

Amplitude

The frequency of the human voice ranges from 85-180 Hz for males and 165 - 255 for females, frequency and pitch are directly connected to the low or high nature of the sound but come from different angles and they are retrieved from the frequency domain analysis while amplitude is gotten from the sampling of the voice.

Database Module

The system at this point takes the audio file and the parameters retrieved from the voice and prepares a search query to the database using the SEARCH Sub Module to find a user whose voice details match the details retrieved from the parameters gotten as a result of the analysis. This database could be any kind from SQL to MS ACCESS and the schema should be structured in such a way that there is a users tables different from the users_voice table where a users can have multiple records on the users_voice table, this allows the system to work better especially due to the fact that a users voice could change because of this mood or emotions can actually affect the loudness of a voice which results in either a high pitch or low pitch.

Output Module

The system returns the result of the search back to the user and the result contains the details of the user whose records match those of the voice which the system just received and if no result is found, the system does something obvious to alert the user that the records did not match any in its database by using the Buzzing sub module or displays the message on the screen using the Display sub module and the user is then given the opportunity to train the system to recognize his/her voice.

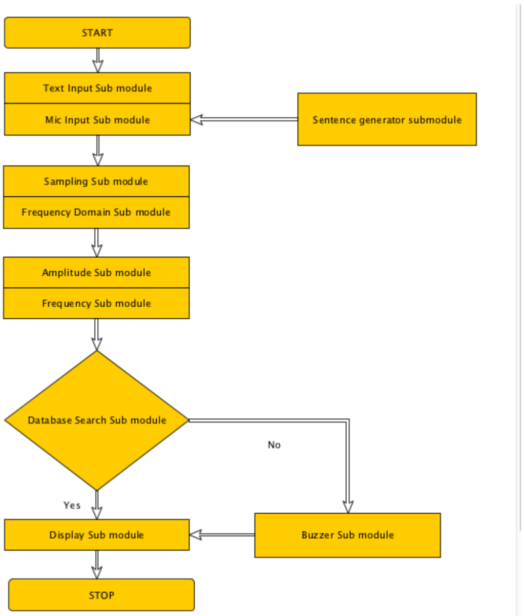


Fig 4. Flow chart representation of the Identification process

TRAINING PROCESS

The system allows users to enroll their voice into the system so that they can be recognized when next they use the system and the training follows the same protocol for finding a users voice however instead of searching for the records it inserts the records into the database so it makes use of the Save sub module when it gets to the Database module.

The system first ask the user to enter some of his/her details such as

Full Name

and Sex

and these details are collected by the Text sub module of the Input Module

After that some words are displayed from the Sentence generator of the General Module which is a public sub module for the user on the screen and the system prompts the user to speak those words, first in a low tone which is recorded then in a high tone which is recorded and then in a mid/moderate tone which is also recorded, the system

then sends the data to the next stage which is the Analysis Module for filtering of noise and analysis of each voice recorded and they are sampled and moved to the frequency domain using the respective Sub modules after which they are sent to the Parameter module where the needed parameters are retrieved from the result of each analysis using the different submodules also.

The system then sends them to the next section which is the Database module and the records are saved to the database, the users name and sex are saved to the users table which has a unique ID then the parameters and the voice file are saved to the users_voice table, each voice file is a database record of its own with a table field called user_id where the unique id of the user in the users table is saved so there has a relationship between the users and the user's voice table which is the id and then a success message is displayed to the user via the Display Sub module of the Output Module.

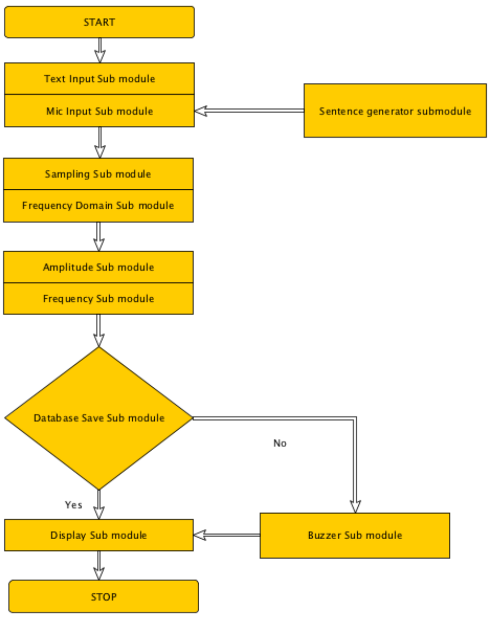


Fig 5. Flow chart representation of the training process

## 3. RESULTS

Results

The Algorithm developed will require certain additional hardware for the system to work such as a microphone, keypad, screen and buzzer, it also needs to have a database such as SQL or MS ACCESS etc for data to be stored, that is data storage for the users data. The system's Speed, Performance will be dependant on the RAM and Processor of the main hardware and accuracy will be depend on the flawlessness of our algorithm.

## 4. CONCLUSIONS

This study has developed a voice recognition algorithm and broken down the system into modules and submodules from input to output and error handling for unforeseen circumstances and it is ok to say the system is efficient, robust and uncomplicated for the recognition of its users voice after training.

## 5. REFERENCES

[1] Imaizumi, S., et al., Vocal identification of speaker and emotion activates different brain regions. Neuroreport, 1997. 8(12): p. 2809-12.

[2] Belin, P., et al., Voice-selective areas in human auditory cortex. Nature, 2000. 403(6767): p. 309-12.

[3] L.A. Streeter et. al. Pitch change during attempted deception. Journal of Personality and Social Psychology, 1977.]