GALGOTIAS UNIVERSITY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

VOICE RECOGNITION USING HMM MODEL- AN UPCOMING TECHNOLOGY

A Report for the Evaluation 3 of Project 2

Submitted by:

RISHABH YADAV

(1613101573/16SCSE101113)

In partial fulfillment for the award of the degree

of

Bachelors of Technology

IN

Computer Science and Engineering

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

Under the supervision of

Dr. NITIN MISHRA, Professor

APRIL/MAY-2020

TABLE OF CONTENTS

ABSTRACT

A voice recognition or speech recognition is a technology which is based on voice. Voice recognition allows a system to create text or open an activity using speech signal. Voice recognition use the process of identification and understanding of speech. It can be use for any language. Voice recognition involves many fields of physiology, psychology, linguistics, computer science and signal processing. The purpose of this project is to deliver natural language communication between man and machine. The speech recognition technology is gradually becoming the key technology of the IT man-machine interface. The paper describes the development of speech recognition technology and its basic principles, methods, reviewed the classification of speech recognition systems and voice recognition technology, analyzed the problems faced by the speech recognition. This technology helps to make our task simpler. Any language drivers can be added in voice recognition. It is beneficial for many field including electronic gadgets.

Keywords- *Voice recognition, Advantages, Hidden markov model, technology*.

INTRODUCTION

Speech recognition system means, an array or system where a computer or any other apparatus is processed by someone's voice by:

- You can procreate calls either by manually choosing the contact or through voice recognition.
- The scheme adapts voice recognition technology to check the person is exactly who they say they are.[1]

Voice recognition doesn't require onscreen or physical keyboard. It is a technological communication. For employers, sanctioning voice recognition in systems and reassuring its use in the workplace can be a 'reasonable adjustment': anticipating discrimination against, and a successful technology for disable users.[2]

To understand the aspect of voice recognition first we need to understand voice acknowledgement. Voice acknowledgment is incorporated with most gadgets where the equipment can bolster it so better quality telephones and tablets will have great amplifiers which will bolster voice input. Thus, PCs regularly accompany inbuilt cameras, amplifiers and speakers. Voice acknowledgment can give an option in contrast to composing on a console. At its easiest, it gives a quick strategy for composing on a PC, tablet or cell phone. The client talks into an outside mouthpiece, headset or inherent amplifier and their words show up as content on the screen. This may be in the content bar of a web crawler, in a talk or delivery person application, or in an email or report. This technology frees people from using a mouse or keyboard and also helps the disabled. It enables a world of productive possibilities.

Voice recognition helps to the potential users. It is acutely useful for a person with a physical disability who finds texting difficult, absurd or impossible.[supra 2] Additionally, it can aid to curtail the risk of getting a constant stress injury or to manage any such upper limb disorder more effectively.

ADVANTAGES OF VOICE RECOGNITION

1. **Useful for disabled person** - There are several disable persons who are not able to use smartphone gadgets. Here the major advantages come. It allows every users to use and work with computer and other gadgets.

2. **Spelling corrections** - Voice recognition is really helpful for people correction mistake and spellings. This technology only use the words which are added in the system. Which means if you don't know the spelling this will automatically correct it for you.

3. **Enhanced speed** - A professional writer can have a speed of 70 - 80 word per minute. But if users use voice for writing a note then can complete their task more quickely. Some people might also get discouraged because of this painful process, therefore to help them speech recognition devices and software are of great help.

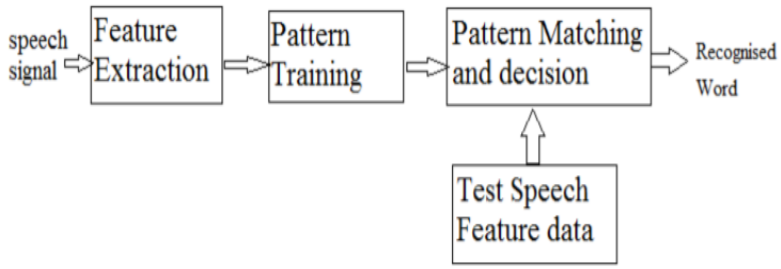
4. **Useful for Language translation -** Voice recognition software can used as a language translate if we add a greater number of language drives. It allows users to operate a computer by speaking to it.

5. **Useful for illiterate person** - Voice recognition software is useful for illiterate person who don’t know how to write words. But an illiterate person is able to speak, therefore is he able to give instructions using is speech.

EXISTING SYSTEM

Speech signal mainly conveys the words or messages being spoken.[3] Ambit of speech recognition deals with identifying the underlying meaning in the revelation.[4] Success in speech recognition depends on obtaining and representing the speech dependent characteristics which can efficiently distinguish one word from another. The speech recognition system may be viewed as working in a four stages as[5]

i. Feature extraction
ii. Pattern training
iii. Pattern Matching
iv. Decision logic



The basic principle of voice recognition involves the fact that speech or words spoken by any human being cause vibrations in air, known as sound waves. These continuous or analog waves are digitized and processed and then decoded to appropriate words and then appropriate sentences.

Generally, Speech recognition system consists of:

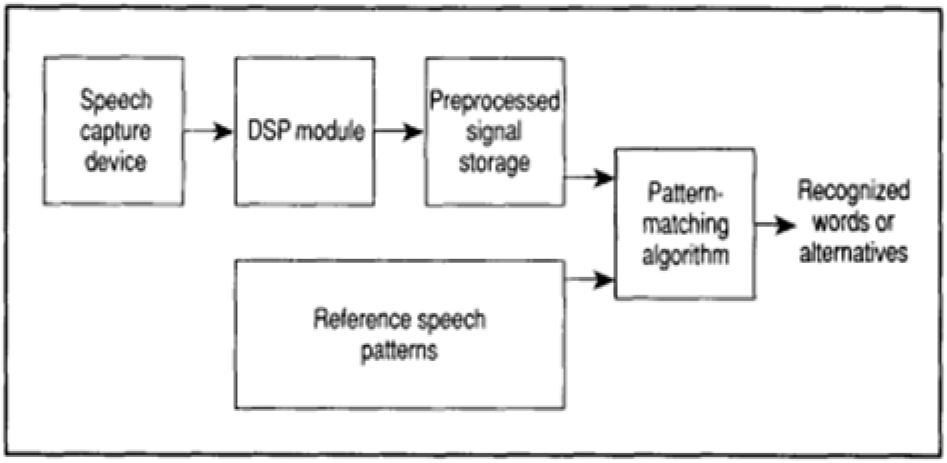
A speech capturing Device: It consists of a microphone, which converts the sound wave signals to electrical signals and an Analog to Digital Converter which samples and digitizes the analog signals to obtain the discrete data that the computer can understand.

A Digital Signal Module or a Processor: It performs processing on the raw speech signal like frequency domain conversion, restoring only the required information etc.

Preprocessed signal storage: The preprocessed speech is stored in the memory to carry out further task of speech recognition.

Reference Speech patterns: The computer or the system consists of predefined speech patterns or templates already stored in the memory, to be used as the reference for matching.

Pattern matching algorithm: The unknown speech signal is compared with the reference speech pattern to determine the actual words or the pattern of words.


Speech capture device
DSP module
Preprocessed signal storage
Pattern-matching algorithm
Recognized words or alternatives
Reference speech patterns

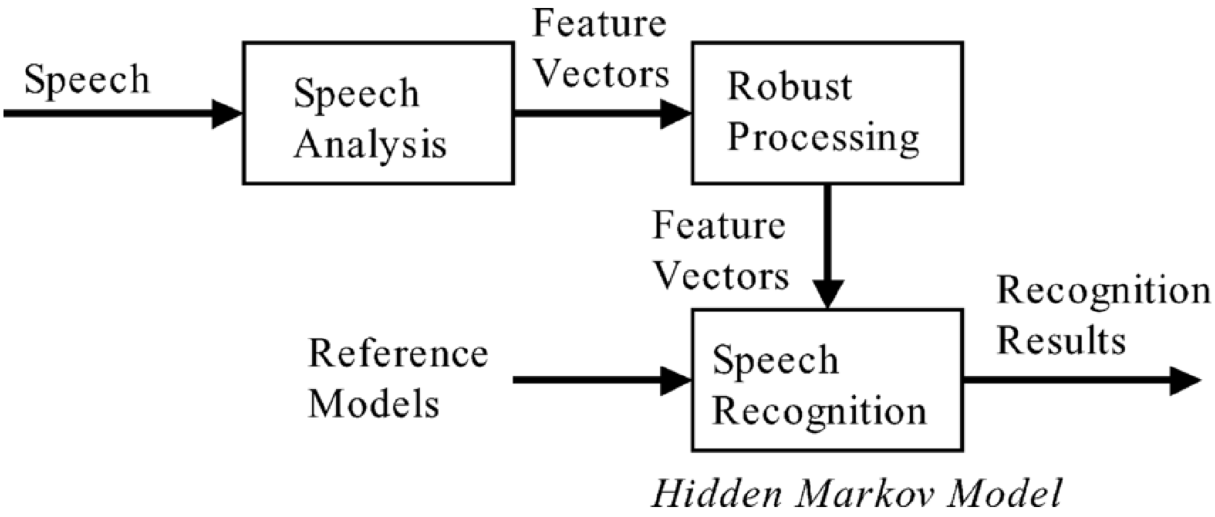
PROPOSED SYSTEM

In the project, there is no novel technological advancement depicted or showed. The present architecture which is best suited for the voice recognition system is described and has been presented. The main objective of the project is to enhance research in regards to the given topic and have in depth knowledge.

The main objective of the voice recognition online research experience was to devise a computer program that distinguishes between speakers. My goal was to create an interdisciplinary project for Computer and Mathematics that applied abstract scientific concepts to a "real life" problem, speaker recognition. The project met the student learning outcome for both majors: "be able to demonstrate the ability to apply mathematical skills to solve computing related problems." A voice recognition problem engaged students to apply the abstract mathematical concept of transforming data using Fourier transforms.

The proposed system in the project showcases the voice recognition system using the HMM Model. Generally the Hidden Markov Model is used. This model develops a non deterministic probability model for the speech recognition. This model consists of two variables – the hidden states of the phonemes stored in the computer memory and the visible frequency segment of the digital signal. Each phoneme has its own probability and the segment is matched with the phoneme according to the probability and the matched phonemes are then collected together to form the correct words according to the stored grammar rules of the language.

IMPLEMENTATION OR ARCHITECTURE DESIGNS



This is how the speech recognition system actually works.

Acoustic model is one of the most important knowledge sources for automatic speech recognition system, which represents acoustic features for phonetic units to be recognized. In building an acoustic model, one fundamental and important issue is choosing of basic modeling units. Generally speaking, when the target language of the speech is specified, there is several types of sub word unit can be used for acoustic modeling. Different acoustic modeling unit can make a dramatic difference on the performance of the speech recognition system. Acoustic modeling of speech typically refers to the process of establishing statistical representations for the feature vector sequences computed from the speech waveform. Hidden Markov Model (HMM) is one of the most commonly used statistical models to build acoustic models. Other acoustic models include segmental models, super segmental models (including hidden dynamic models), neural networks, maximum entropy models, and (hidden) conditional random fields, etc. An acoustic model is a file that contains statistical representations of each of the distinct sounds that makes up a word. Each of these statistical representations is assigned a label called a phoneme acoustic model is created by taking a large database of speech called a speech corpus and using special training algorithms to create statistical representations for each phoneme in a language. Each phoneme has its own HMM. The speech decoder listens for the distinct sounds spoken by a user and then looks for a matching HMM in the acoustic model. Each spoken word w is decomposed into a sequence of basic sounds called base phones. The acoustic model describes the probability of a specific observation given a base phone.

1. Speech Input

Speech input is taken from a microphone attached to the system sound card, the sound card handles the conversion Of the analogue speech signal into digital format. Depending on the recognition software type i.e. Continuous or non-continuous the engine may need to listen continuously for sound input, in this case a continuous audio stream will need to be opened and read from (Section 3 documents our implementation Of this procedure using the Java Sound APO.

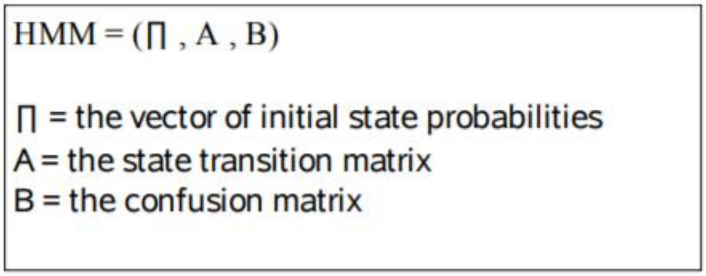
2. Signal Processing

Speech input is taken from a microphone attached to the system sound card, the sound card handles the conversion Of the analogue speech signal into digital format. At this point we have a digitised version Of the speech signal. Speaking comes naturally to people, when we are spoken to we hear individual words, sentences and pauses in the speech, more so our understanding Of language allows to interpret what was said. Consider what happens when we hear people speaking in a language which is foreign to us, we don't hear the individual words in the language and the speech sounds like one continuous stream Of noise. The same scenario is true when we speak to computers for the purposes Of speech recognition. The process of finding word boundaries is called segmentation.

Applying this knowledge to our digitised signal data, we need to process this signal in order to determine the word boundaries in the Signal and also to extract.

3. Hidden Markov Models

Hidden Markov Models (HMM) have proven to date to be the most accurate means of decoding a speech signal for recognition. HMM are stochastic in nature, that is, they generate a probability that an outcome will occur. In our speech system the input to the HMM will be a speech signal sampled at a particular moment in time, the output from the Markov Model will be a probability that the inputted signal is a good match for a particular phoneme. We can create numerous Hidden Markov Models to model speech signal input at time samples of the signal, each Markov Model can represent a particular phoneme. By combining the probabilities from each Markov model we can produce a probabilistic output that a given speech sequence is a representation of a particular word.



HMM = (∏ , A , B)

∏ = the vector of initial state probabilities
A = the state transition matrix
B = the confusion matrix

The fundamental concept in a HMM is the Markov assumption, this concept assumes that the state of the model depends only on the previous states. Given this concept, we can introduce the Markov process, a process which moves from state to state depending only on the previous n states.

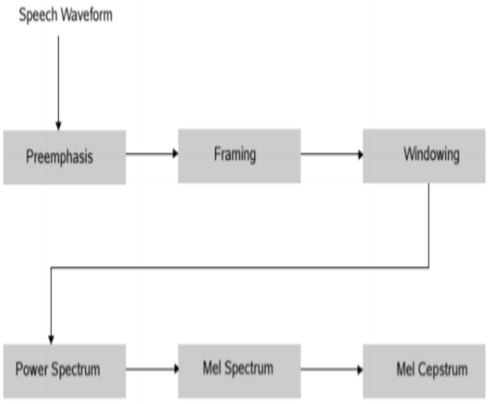
A represents the transition matrix or the probability of transiting from one state to another. B represents the confusion matrix or the set of observation likelihood's, which represent the probability of an observation being generated from a given state. Having modelled the speech input utterances using Hidden Markov Models we need to determine which observation sequence produces the best probability, we use the Viterbi algorithm to do this. The goal of the Viterbi algorithm is to find the best state sequence q given the set of observed phones.

4. Grammar

Given a speech corpus which consists Of many thousands Of speech samples, during the recognition process the hidden markov model may have to search the entire corpus before finding a match. This searching problem is known as perplexity and can cause the speech engine to operate inefficiently, which translate into delays in recognition for the user. Grammars can be used to reduce perplexity. Grouping Markov Models Of phonemes together we can form words. For example, a grammar can specify a particular word order, such that if we can recognise word A and we know that word B never follows word A we can eliminate searching the corpus for word B.

5. Recognised Speech

The output from these steps of the process is a String of text which represents the spoken input, at this point the recognised speech can be applied to some application domain such as in a dictation systems or, in our case, a command and control scenario.



The Hidden Markov model is implemented as the java object in this technology. The constructor of the Markov takes two integers corresponding to the number Of states and number Of observation symbols. The initial state is set to I and the transition probabilities are initialised to random values. The Viterbi find the best (most probable) path through the Markov trellis. The input to the algorithm is an observation sequence corresponding to an input signal (i.e. speech utterance). The signal has been pre-processed by the feature extraction stage; the Viterbi algorithm returns the probability that the input utterance is a recognised word. The training process for the Hidden Markov Model involves the use Of the Baum Welch Algorithm. This algorithm is designed to find the HMM parameters.

OUTPUT/RESULT

Automatic speech recognition (ASR) system is a hardware and software system, where the input is the sound of the voice (speech) and the output is the identification of those spoken words.

The project gives a complete understanding of the working of the HMM Model in the speech recognition system. It clearly the implementation of HMM model using the Java object.

Acquisition Of the speech signal is achieved through a microphone using the Java Sound APL. The audio signal is recorded as pulse code modulation (PCM) with a sample rate Of 16KHz, this is implemented in Java using floating point numbers. The Java Sound API objects Target Data Line and Audio Format are used to create the input data line from the sound card, the method open() called on the Target Data Line Object opens the line for audio input. The Audio Format Object is used to create audio input Of the specified type, in our case this is PCM signed with a frequency Of 16KHz. Code fragment I shows the basic Java code used to open a line for audio input. The Java code in this class is implemented as a thread, which allows the system to do other work in the recognition process while continuously listening for audio input.

CONCLUSION/FUTURE WORK

For a suitable development platform java is a best programming language which is capable of APIs. These API's are use to develop the technology of voice recognition. How the Java sound API can be used to develop sound recognition we have seen here. The basic speech recognition engine is at place, to deploy it on ipaq and Linux the next phase is used.

The current implementation of this system requires the user to learn words, phrases and languages. The future work of voice recognition is to learn every new word and language which are present all around the world. Not only that the enunciation should be improve and it should be developed more fluent. Training of voice is required which it can learn after real time use. Integrating a speaker independent component into this system through a speech corpus, would be the next logical step in the system evolution.

REFERENCES

[1] Cambridge Business English Dictionary, (April 13, 2020) (https://dictionary.cambridge.org/dictionary/english/voice-recognition)

[2] Voice recognition- An overview

[3] Vrushali Bhamare, Priyanka Kalokhe, Ketaki Kulkarni, ”Grammatically Correct Speech by Using Speech Recognition”: International Journal of Innovative Research in Computer Science and Communication Engineering.[ISSN No.-2320-9801]

[4]Rabiner and Juang, ”Information Science and Computing” International Book Series, 1983

[5] Ms. Rupali Chavan, Dr. Ganesh. S Sable, “An Overview of Speech Recognition Using HMM” : International Journal of Computer Science and Mobile Computing.[ISSN No.- 2320-088X]

[6] Qiaohong Zu Bo Hu(Eds.),”Human Centered Computing”; Second International Conference, HCC, Colombo, Sri Lanka, January 7-9,2016.

[7] Kanchan Naithani, V.M. Thakkar, “English Language Speech Recognition Using MFCC and HMM”: International Conference on Research in Intelligent and Computing in Engineering (RICE)

[8] Nils Bagge, Chris Donica, “Text Independent Speaker Recognition” ELEC

[9] R. EJBALI, Y. Benayed, M. ZAIED and A. ALIMI REGIM, “Wavelet Networks for phonemes Recognition”: International Conference on systems and processing units”.