

# A Novel Speech Recognition System using Hidden Markov Model

R. Renita Raxy<sup>1</sup>, R. Rosita Rosini<sup>2</sup> and G. Mani Sankar<sup>2</sup>

<sup>1</sup>Department of ECE, Bharath University, Selaiyur, Chennai – 600073, India;

<sup>2</sup>Department of ECE, Kalasalingam Institute of Technology, Anna University, Krishnankoil - 626126, India

## Abstract

Here in this paper, we have made a formal assessment of the current capabilities of Speech Recognition system and have taken efforts to show how they have been exploited in present day applications using Hidden Markov Model (HMM). Voice Quest is a huge leap in the world of digital communication. The objective of this paper is to analyze and explain the importance of Voice Quest. Voice quest aims to make searching more easy and simple. First of all, we have to create a database that can accommodate certain questions and their appropriate answers. Whenever user asks a question, this input question is first converted to its text format and tries to get a particular match from our database. After getting that match, the corresponding answer for that question, which is stored in the database in text format is converted to voice format and is given as the output to the user. Since the Voice Quest avoids typing and button presses, it makes searching quite a simpler task.

**Keywords:** Digital Communication, HMM, Speech Recognition, Voice Quest

## 1. Introduction

Speech is no doubt the most powerful medium of human interaction. By means of modern digital signal processing, human-machine interaction has also come to a reality. There are a number of reasons to justify why users wish to speak and listen to mobile devices. We learn how to speak and listen to other people in the first couple of years of our lives, so it is quite natural that gradually we tempt to speak and listen to our mobile devices also.

This paper is about Voice Quest. Voice Quest is an enhancement of Speech Recognition and Hidden Markov Model (HMM) has been used for the purpose. In this paper we are trying to make searching as easy as possible. And we think, Voice Quest is the best way because without much effort it can provide an appropriate answer to a question of the user. One of the main objectives of Voice Quest is to help people with physical disabilities. It can be of great help to people who find typing difficult, painful

or impossible. Voice Quest is a God-Send for people with spelling difficulties. Voice Quest plays a noticeable role in education by helping people with learning disabilities. Reading and spelling words is a big challenge faced by people suffering from dyslexic. For such people, Voice Quest is going to be a great blessing. Voice Quest will be a remarkable achievement for illiterate and visually challenged people.

The usage of mobile devices is increasing day by day and everyone is waiting for new developments and applications that can make life much easier. Voice is a much comfortable way of interacting with a mobile device, while driving a car or operating complex machines. Of course, it is going to be a time saving and useful application for its users. In today's busy life, using your voice as a means to search can be much quicker and easier than typing<sup>1</sup>. Here in this paper using Voice Quest, user can search for the answer of a particular question. For example, if the user asks a question like, "Name the

\* Author for correspondence

person who have written the national anthem?", then the device will return the appropriate answer to the user in voice format itself. The user gets an answer to his question without any typing or button-presses. The introduction of such a time saving application is essential to beat today's fast life.

Voice Quest would provide a much user friendly and natural way of searching because is just like asking a question to one of your friends. Here for searching an answer to a particular question the user just need to ask the question to the device which is in front of him and the device is ready with an answer. Since no typing or button-presses is required the users hand is left free to do some other task in the meantime. Voice Quest is a growing technology that may gradually change the interface between humans and machines. The current interface between humans and machine is the traditional mouse and keyboard. This will definitely be an added advantage for the busy users of the present day world.

## 2. Existing System

This research work is based on the existing Speech Recognition system. Speech Recognition was introduced to talk to machines. It is truly an alternative to typing. With this technology, you can simply talk to a machine and the words uttered by the user will appear on the screen<sup>2</sup>. Voice to text conversion can be achieved by executing the following 3 steps<sup>3</sup>.

- Converting Audio
- Breaking Speech into Segments
- Conversion of Speech Segments

### 2.1.1 Converting Audio

The first step converts the words spoken by the user to digital data. The idea that is conveyed by the users in words is converted to a digital stage. Basically, a user use devices like the microphone to convert his voice to an analog signal. The initial step in this process begins by converting voice to a machine understandable language/machine readable input<sup>4</sup>. When user speaks to the machine, vibrations are created in the air. An analog-to-digital converter is used to grasp these vibrations and this converter will translate these analog waves to digital data that the machine can understand.

This analog to digital converter will digitizes or computerizes the idea conveyed by the user through

voice by attaining accurate estimations of the vibrations at persistent intervening time. This digitized idea that is conveyed by the user is then filtered to take away all sorts of background noises that are not desired.

### 2.1.2 Breaking Speech into Segments

Now the words spoken by the user is turned from analog input to digital input. In this step, the digital input entered for processing is reduced into smaller units. Once the digital input is reduced to smaller units, these units are examined in order to find a match with the known phonemes. A phoneme is the principle foundation and the smallest bit of a language's length of utterance or phonology. Phonemes when combined with other phonemes we get words and sentences which are meaningful. To illustrate this general concept, consider the following example, 'p' and 'b' in the English words 'park' and 'badge'. Pretty nearly there are 40 phonemes in our English language. After digitizing our analog input, the machine will convert these phonemes into words having a meaning. Then the machine will convert these words to pretty good sentences that can convey users thought. Each phoneme has got a single underlying or fundamental frequency and this distinctive aspect makes the identification of every single phoneme at recognition stage. Usually speech is recognized by comparing the input speech with the stored templates of phonemes. And the most matching phoneme or word is given as output. In this phase, every single syllable of the language spoken by the user is identified and is converted to text.

### 2.1.3 Conversions of Speech Segments

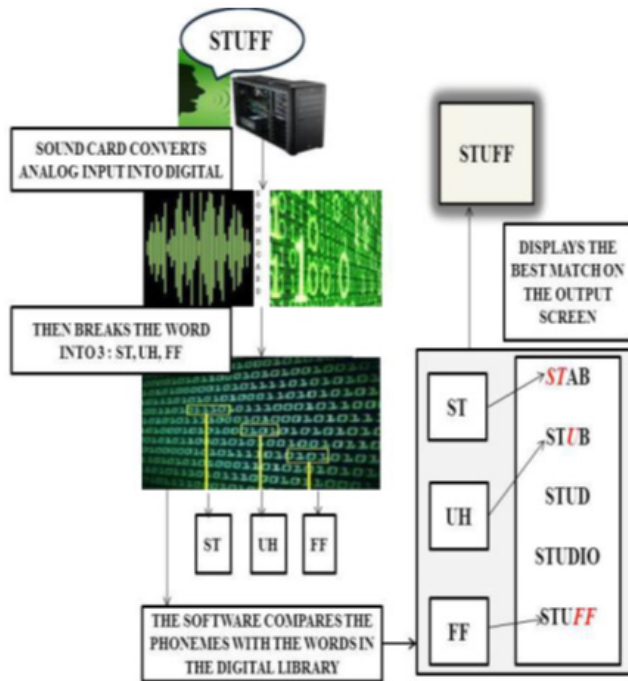
In the finishing stage of this process the words uttered by the user get converted into the digital forms. And it's the most difficult part in this process. Here different phonemes are combined together and this combination of phonemes is compared with a digital library of well-known words and sentences.

## 2.2 Hidden Markov Model

The Hidden Markov Model (HMM) is a finite set of states. Each of which is related with a probability distribution. Transitions among the states are controlled by a set of probabilities. These sets of probabilities that are responsible for controlling the transitions among the different states are called transition probabilities. Speech

Recognition can be put into effect with the help of HMM. HMM consists of the following four steps.

- Training.
- HMM-Based Recognition.
- Digit Models.
- Text Storage



**Figure 1.** Conversion of Speech to Text.

### 2.2.1 Training

Training is a highly significant part of speech-to-text conversion. It encompasses producing a pattern representative of the property of a class by applying one or more test patterns that have close similarity to speech sounds of the identical class.

### 2.2.2 HMM-Based Recognition

It is a particular procedure for comparing the undisclosed test pattern with each sound class reference pattern and computing a measure of similitude between the test pattern and each reference pattern.

### 2.2.3 Digit Models

Pre-process the input speech sample, and extracts the feature vector. After that the index of the nearest codebook vector for each frame is sent to all digit models. The model having the highest possible probability is taken

as the recognized digit.

### 2.2.4 Text Storage

The digital speech data is forwarded to a PC. Then the target program which is running on the PC will collect the text and record it to the disk.

## 3. Proposed System

This research work is based on the existing Speech Recognition system. As we all know with the introduction of Speech Recognition we were able to talk to machines. The main focus behind this article is to make this digital communication more natural. Technology is developing, Life is changing, Needs and demands of people are increasing day by day. Therefore, talking to machines is not a fantasy anymore.

As mentioned earlier, our proposed system will respond to users query within no time. To make this possible first we have to create a database to hold a couple of questions and their appropriate answers. Here we are limiting our database to store only General Knowledge questions. Now when a user asks a question to the machine it will first convert the users spoken words into text by using Speech Recognition. Then it will check its database to find a suitable match. After getting a suitable match from the database it will fetch the answer for that particular question and will convert the answer from text to voice format using text to voice conversion.

Here we would like to present a brief description about text to voice conversion.

In text to voice conversion the machine reads out each and every single distinct meaningful element of speech to the user. Machines may use devices like loudspeaker for this purpose<sup>5</sup>.

There are three essential stages involved in this process

- Conversion of text to words
- Conversion of words to phonemes
- Conversion of phonemes to sound

### 3.1 Conversion of Text to Words

Reading out every single distinct meaningful element of a text is not easy as it sounds. The main problem is about the ambiguity of the written text. A written input may have a different meaning depending upon the different context in which it is used. So, here the meaning of the written input must be grasped correctly or an intelligent guess has

to be made to read out the input words correctly without altering the idea that the user was trying to convey.

Therefore the first step in speech synthesis is speech pre-processing. Speech pre-processing is all about bringing the ambiguity or inexactness to a minimum extent. It is all about shrinking down the many distinct styles you could read an instance of a text, into the best way to convey the user's idea. Speech pre-processing proceeds through the text and clean the unwanted background noises, so that the machine makes hardly any mistakes when it actually reads the text audibly.

### 3.2 Conversion of Words to Phonemes

Having reached a conclusion about the words that need to be read out, the speech synthesizer is now supposed to give rise to the sounds that generate those words. Machine need to have a large digital library to store the list of these words in its alphabetical order. And it also stores a list of phonemes associated with each of these words to produce its sound.

### 3.3 Conversion of Phonemes to Sound

So now the sequence of the input text is converted to its corresponding list of phonemes. Normally we have three different approaches to store these phonemes in our digital library.

- One is to use recordings of a person saying the phonemes.
- By generating basic sound frequencies, machines generate phonemes by their own.
- Imitate the mechanism of the human voice.

### 3.4 Advantages of our Proposed System:

It avoids manual typing to search the answer for a question. Therefore we can say that it is a better choice.

It is more efficient because when a particular question is asked to the machine it speaks out the appropriate answer rather than giving a list of links as in the traditional searching. Make the user's task much easier.

It is a time saving application. Here when a question is asked to the machine it will check whether it is there in its database or not. If user's question is present in its database, the machine will immediately respond to the user by telling the appropriate answer.

It can assist physically challenged people to a great

extent to lead a better life without depending much on other individuals.

This article is a promise to our busy future generation to lead a better life.

Android is one of the most powerful and flexible operating system. And that is the main reason for which it is preferred by most of the users. Moreover Android presents a user friendly platform for its users. These positive features of Android have motivated us to prefer an Android platform to implement our paper.

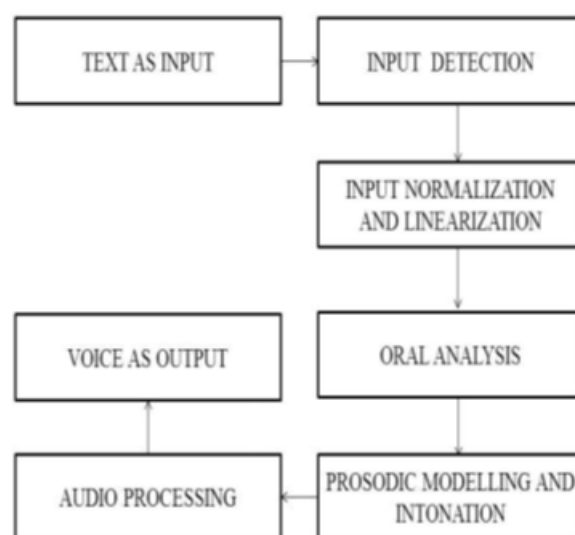


Figure 2. Conversion of a Text to Voice.

## 4. Conclusion

Hope this article gave you a decent knowledge about how Voice Quest works. The main aim behind this paper is to bring out a change in the traditional searching technique and also to make searching more user friendly. Now the technology is growing day by day so as the needs of the people. As the needs and demands of the people are increasing, we hope such an application is going to hit this society with an unimaginably positive impact.

Voice Quest enables a user to search the answer for a particular question without the traditional human-machine interface like keyboard and mouse.

This application can be of a great help to people with disabilities and also for people who are engaged in multiple tasks. Voice Quest has a good future in educational field also, because it can help students with learning disabilities

a lot. In the coming years we will definitely witness a new era of digital communication.

## 5. Future Work

The Voice Quest system introduced in this paper has certain drawbacks which can be corrected in future as technology improves.

In this paper we have made an attempt to provide a review of how much Speech Recognition has progressed in the previous years and an enhancement of the same has been introduced.

We are continuously struggling to make remarkable achievements in this field. Further we are aiming at following improvements.

One of the main limitations of our proposed work is that it is limited to a single language (i.e. only English). In the coming years efforts can be taken to improve Voice Quest by including different languages.

In this paper we have limited our database to store only General Knowledge questions. In future this application can be improved by expanding database to hold several questions from different fields.

## 6. References

1. Kimio T, Natarajan G, Hideki A, Taichi K, Nanao K. Higher involvement of subtelomere regions for chromosome rearrangements in leukemia and lymphoma and in irradiated leukemic cell line. *Indian Journal of Science and Technology*. 2012 April; 5(1):1801–11.
2. Cunningham CH. *A Laboratory Guide in Virology*. 6th ed. Minnesota: Burgess Publication Company; 1973.
3. Sathish Kumar E, Varatharajan M. Microbiology of Indian Desert. In: Sen DN, editor. *Ecology and Vegetation of Indian Desert*. India: Agro Botanical Publishers; 1990. p. 83–105.
4. Varatharajan M, Rao BS, Anjaria KB, Unny VKP, Thyagarajan S. Radiotoxicity of Sulfur-35. *Proceedings of 10th NSRP; India*. 1993. p. 257–8.
5. Karpe V, Kabadi S, Mutgekar A, Pokharkar M. Controlling device through speech recognition system. *International Journal of Advanced Research in Computer Science and Software Engineering*. 2014; 4(2):1020–4.
6. Raj MS, Saravanan T, Srinivasan V. A modified direct torque control of induction motor using space vector modulation technique. *Middle - East Journal of Scientific Research*. 2014; 20(11):1572–4. ISSN: 1990-9233.
7. Reddy BR, Mahender E. Speech to text conversion using Android platform. *International Journal of Engineering Research and Applications (IJERA)*. 2013; 3(1):253–8.
8. Bharatwaj RS, Vijaya K, Rajaram P. A descriptive study of knowledge, attitude and practice with regard to voluntary blood donation among medical undergraduate students in Pondicherry, India. *Journal of Clinical and Diagnostic Research*. 2012; 6(S4):602–4. ISSN : 0973 - 709X.
9. Kaur HS, Chaudhary V. Literature survey on automatic speech recognition system. *International Journal of Advanced Research in Computer Science and Software Engineering*. 2014; 4(7):398–402.
10. Anbuselvi S, Rebecca J. A comparative study on the biodegradation of coir waste by three different species of Marine cyanobacteria. *Journal of Applied Sciences Research*. 2009; 5(12):2369–74. ISSN : 1815-932x.
11. Singh PP, Singh EB. Speech Recognition as emerging revolutionary technology. *International Journal of Advanced Research in Computer Science and Software Engineering*. 2012; 2(10):410–3.
12. Rajasulochana P, Krishnamoorthy P, Dhamotharan R. An investigation on the evaluation of heavy metals in *Kappaphycus alvarezii*. *Journal of Chemical and Pharmaceutical Research*. 2012; 4(6): 3224–8. ISSN : 0975 – 7384.
13. Jose DV, Mustafa A, Sharan R. A novel model for speech to text conversion. *International Refereed Journal of Engineering and Science (IRJES)*. 2014 Jan; 3(1):39–41.
14. Jasmine MIF, Yezdani AA, Tajir F, Venu RM. Analysis of stress in bone and microimplants during en-masse retraction of maxillary and mandibular anterior teeth with different insertion angulations: A 3-dimensional finite element analysis study. *American Journal of Orthodontics and Dentofacial Orthopedics*. 2012; 141(1) :71–80. ISSN : 0889-5406.