

# Acoustic Authentication Techniques and Applications: A Review

Astha Gupta  
Assistant-Professor

Department of Information Technology  
Chandigarh Engineering College

Neeraj Sharma  
Assistant-Professor

Department of Information Technology  
Chandigarh Engineering College

**Abstract:** The earliest advances in speech recognition centered in the main on the creation of vowel sounds, because the basis of a system that may conjointly learn to interpret phonemes (the building blocks of speech) from close interlocutors. [1] These inventors were hampered by the technological context during which they lived, with solely basic suggests that their disposal to create a talking machine. Even so, they supply necessary background to more modern innovations. The history of the technology reveals that speech recognition is way from a brand new preoccupation, albeit the pace of development has not forever matched the amount of interested within the topic. The appeal is speech recognition is instinctive and fascinated that machines can understand us. In this paper, we take a trip through the voice recognition technology like Dynamic Time Wrapping, Hidden Markov Model, Artificial Neural Network and Vector Quantization used in the past and their comparison and applications.

**Index Terms:** Speech Recognition, Dynamic Time Wrapping, Hidden Markov Model, Artificial Neural Network and Vector Quantization.

## I. INTRODUCTION

The fundamental definition for speech must be founded on voice. Voice (or vocalization) is characterized as the sound delivered by people and different vertebrates utilizing the lungs and the vocal creases in the larynx, or voice box. Voice isn't constantly delivered as speech, be that as it may. Babies jibber jabber and coo; creatures bark, moo, whinny, snarl, and howl; and grown-up people chuckle, sing, and cry. Voice is produced via wind current from the lungs as the vocal folds are brought close together. When air is pushed past the vocal folds with adequate weight, the vocal folds vibrate. In the event that the vocal creases in the larynx did not vibrate, discourse must be delivered as a murmur. The voice is as one of a kind as unique mark. [1] It characterizes identity, inclination, and wellbeing. Speech is the voice conveying a thought with the assistance of a dialect.

### A. Voice Recognition

Since ages, the only essential mean of communication between people is the Voice. Voice Recognition is the way toward changing over an acoustic speech into content, and/or recognizable proof of the speaker. Throughout the years with late appearance in innovation it has turned into a basic and basic piece of our way of life because of the expanding correspondence among human and PCs or mechanized frameworks.

Voice or speaker recognition is the capacity of a machine or program to get and translate transcription or to comprehend and do spoken commands. Voice recognition has picked up noticeable quality and use with the ascent of AI and keen associates.[3] Voice recognition frameworks empower purchasers to interface with innovation essentially by addressing it, empowering without hands solicitations, updates and other straightforward undertakings.

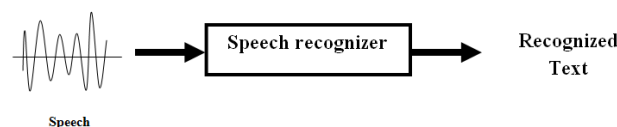


Fig.1 Voice Recognition System

The Voice may be a signal of infinite information. Digital process of speech signal is incredibly vital for top speed and precise automatic voice recognition technology. These days it's getting used for health care, telecommunication military and other people with disabilities thus the digital signal processes like Feature Extraction and have Matching are the newest problems for study of voice signal. So as to extract valuable info from the speech signal, build selections on the method, and procure results, the info must be manipulated and analyzed. Basic technique used for extracting the options of the voice signal is to seek out the Mel frequency cepstral coefficients. Mel-frequency cepstral coefficients (MFCCs) are the coefficients that conjointly represent the short power spectrum of a sound, supported a linear circular function remodel of a log power spectrum on a non-linear mel scale of frequency. [4]

Voice recognition permits you to supply input to an application with your voice. A bit like clicking with mouse, writing on the keyboard, or pressing a key on the phone input device provides input to voice application, voice recognition system give input by talking within the desktop world, a mike is needed to do that. In the Voice XML world, all you need a phone. The voice recognition method is performed by a code part referred to as the speech recognition engine. The primary function of the voice recognition engine is to method spoken input and translate it into text that an application understands. Once the user says something, this is referred to as utterance. Utterance is any stream of speech between two periods of silence. Utterances are sent to the speech engine to be processed then human voice is regenerate into digital signal kind to provide digital knowledge representing every level of signal at each separate time step.

## B. Block Diagram for Voice recognition

The basic architecture for the recognition of voice can be interpreted like the figure mentioned below. The Speech signal feature is extracted, on which the recognition is done, i.e. the extracted features are matched with the features already saved in the database (patterns) and then, the voice is recognized.

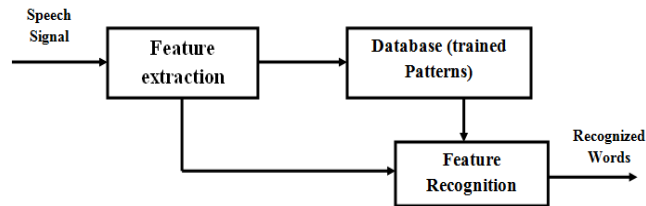


Fig.2 Block Diagram of Voice Recognition System

## II. TECHNIQUES USED IN THE PAST

The Voice Recognition framework is basically an example recognition framework, including feature extraction, design coordinating, and the reference model library. The obscure voice through the amplifier is changed into an electrical flag on the contribution of the recognizable proof framework, the first after the pre-treatment. The framework builds up a voice demonstrate as per the human voice attributes, dissects the information voice flag and concentrates the required highlights on this premise, it sets up the required format of the discourse acknowledgment. PC is utilized in the recognition procedure as per the model of Voice Recognition to analyze the voice layout put away in the PC and the qualities of the info voice flag. Voice Recognition techniques incorporate dynamic time warping (DTW), hidden Markov model (HMM), vector quantization (VQ), artificial neural network (ANN), support vector machine (SVM).

### A. Dynamic Time Wrapping

DTW is a strategy that ascertains an ideal match between two given groupings (for example time arrangement) with certain confinement. Dynamic time Wrapping (DTW) is one of the calculations for estimating closeness between two fleeting successions, which may fluctuate in speed. For example, likenesses in strolling could be recognized utilizing DTW, regardless of whether one individual was strolling quicker than the other, or if there were increasing speeds and decelerations over the span of a perception. DTW has been connected to worldly successions of video, sound, and designs information — in reality, any information that can be transformed into a straight grouping can be dissected with DTW. [5]

### B. Hidden Markov Model

A Hidden Markov Model (HMM) is a measurable model in which the framework being displayed is thought to be a Markov procedure with obscure parameters; the test is to decide the Hidden parameters from the recognizable information.

In a Hidden Markov Mode, the state isn't specifically noticeable, yet factors affected by the state are unmistakable. Each state has a likelihood conveyance over the conceivable yield tokens. Consequently the arrangement of tokens produced by a HMM gives some data about the succession of states. A Hidden Markov Model can be viewed as a speculation of a blend demonstrate where the shrouded factors which control the blend segment to be chosen for every perception, are connected through a Markov procedure instead of free of one another.

### C. Vector Quantization

Vector Quantization (VQ) is a strategy in which the mapping of vector is performed from an expansive vector space to a limited number of locale in that space. This procedure depends on square coding guideline. Every area is called as group and can be spoken to by its inside known as a code-word. Code book is the accumulation of all code-words. [6]

### D. Artificial Neural Network

Artificial Neural Network (ANN) are only the unrefined electronic models dependent on neural structure of mind. The human cerebrum fundamentally gains from the encounters. Some issues which are past the extent of current PCs can be are effortlessly resolvable by vitality effective bundles.[3] Such sort of cerebrum demonstrating likewise gives a less specialized way to the improvement of machine arrangement. ANN are PC having their design displayed after the mind. They

fundamentally include several straightforward preparing units wired together in complex correspondence arrange. Every straightforward preparing unit speaks to a genuine neuron which sends off another flag or fires on the off chance that it gets a solid flag from the other associated unit.[6] Further there are different types of Neural Network Feedforward Network, Recurrent Neural Network, Modular Neural Network, Kohonen Self Organizing Maps.

### III. COMAPRISON OF DIFFERENT VOICE RECOGNITION TECHNIQUES

Table 1 Comparison of different Voice Recognition Techniques

Techniques Name	Advantage	Limitation
Dynamic Time Wrapping	<ul style="list-style-type: none"> <li>• Its implementation is well-studied and relatively easy.</li> <li>• Reasonable results for small vocabularies[8]</li> </ul>	<ul style="list-style-type: none"> <li>• Deal with uncertainty.</li> <li>• Difficult to generalize the DTW system to continuous speech.</li> </ul>
Hidden Markov Model	The most successful approach to achieve accurate speech recognition system.[8]	Need very sophisticated calculations and a longer training data sequence
Vector Quantization	It has the ability to develop new ideas in combination with other techniques like ANN.	The designing of code book in VQ is the difficult part.
Artificial Neural Network	<ul style="list-style-type: none"> <li>• It is a nonlinear model which is easy to use and understand.</li> <li>• Are flexible in changing environments.[8]</li> </ul>	<ul style="list-style-type: none"> <li>• No structured methodology.</li> <li>• May give unpredictable output quality.</li> <li>• Empirical nature for model development.</li> </ul>

#### IV. APPLICATION OF VOICE RECOGNITION

Voice Recognition frameworks had discovered an expansive scope of utilization in modernized recreations and toys, control of various instruments, information gathering, and transcription. The component likewise turned out to be of much help among the individuals who couldn't get keypads and among those with specific disabilities. Siri, which is introduced on the most recent iPhones, is among the most noticeable case of portable voice interface and it demonstrates the effect of Voice Recognition in the present society. The accompanying outlines a portion of the effects of Voice Recognition in the society.

##### A. Used in evolving search engine:

When utilizing web crawlers there can be contrasts between how we type our request and how we verbalize similar inquiries. The client may experience difficulty communicating an expression or their goal subsequently may not gain suitable outcomes.[4] With the consideration of Voice Recognition in web indexes, the outcomes exactnesses will be essentially expanded. As Voice Recognition enhances, there will be a critical ramifications on how people in general perspectives web search tools by and large.

##### B. Impact in the healthcare industry:

The component has its utilization in medical reporting by medical personnel. When it was presented in this industry doctors were unable to achieve tasks. The framework had a restricted comprehension understanding of medical terminologies. Along these lines, specialists needed to learn on the most proficient method to converse with the product. The innovation was enhanced to be easy to use and precise; this was built up by basic upgrades and consideration of important vocabularies.

##### C. Use in service delivery:

Clients and customers might not have any desire to address a live administrator. Therefore, they pick to utilize the voice recognition frameworks. This makes the procedure productive and enhances time as it cuts on holding up time. This has its application in different air terminals in affirming travel timetables of the air ship.

##### D. Automated identification:

In order to abstain from giving delicate and dangerous individual data, foundations may select to utilize speech recognition to validate personalities of their customers. This has controlled misrepresentation and telephone wrong doings by utilization of voice biometrics in specific organizations like banks.[9]

##### E. Communication in service providers:

Media transmission suppliers use speech recognition to serve their customers who might need to get client care administrations. This comprises of different inquiries by the product to set up the guest's requests and afterward guides them to the suitable administrator for help.

#### V. RESEARCH GAP & PROBLEM FOUNDATION

At present, speech recognition inquire about advancement has been moderate; mainly in theory but has no leap forward. In spite of this fact, new things keep on developing, yet in addition the absence of general materialness. For the most part in:

1. Poor versatility of the speech recognition framework is essentially thought about in the reliance the environment, on the off chance that you gathered speech preparing framework in specific conditions, the framework must be application in this condition; generally the framework execution will be a sharp decrease.
2. It does not react accurately for the error contribution of clients.
3. Additionally, the advancement of speech recognition in uproarious conditions is exceptionally troublesome, in light of the fact that as of now individuals articulate fluctuates significantly, similar to voice, moderate voice rate, pitch and formant changes, which is the Lombard impact, must locate another flag investigation and preparing approach.
4. Understanding of the human sound-related perception, the aggregation of information and learning component and arrangement of the mind control system is as yet hazy.

#### VI. CONCLUSION

Today after extreme research, Voice Recognition System, have made a specialty for themselves and can be seen in numerous different backgrounds. The precision of Speech Recognition Systems stays a standout amongst the most critical research difficulties for example noise, speaker variability, language variability, vocabulary size and domain. The plan of Voice Recognition framework requires watchful considerations to the difficulties, for example, different kinds of Speech Classes and Speech Representation, Speech Pre handling stages, Feature Extraction systems, Database and Performance assessment.

Voice Recognition and distinguishing proof innovation centers around preparing the framework to perceive a person's one of a kind voice qualities (i.e., their voice print). The innovation loans itself well to an assortment of employments and applications, including security get to control for mobile phones (to take out PDA extortion), ATM producers (to dispense with stick # misrepresentation) and vehicle makers (to significantly lessen robbery and carjacking). To have the capacity to control gadgets by voice has dependably fascinated humanity.

## REFERENCES

1. Nitin Washani and Sandeep Sharma, "Speech Recognition System: A Review", International Journal of Computer Applications, Volume 115 – No. 18, April 2015.
2. Jackson Zhang and Bruce Wang, "A Novel Voice Recognition model based on HMM and fuzzy PPM", IEEE, 2010.
3. Ganesh K Venayagamoorthy, Viresh Moonasar, Kumbes Sandrasegaran, "Voice Recognition using Neural networks", IEEE, 1998.
4. Jianliang Meng, Junwei Zhang, Haoquan Zhao, "Overview of the Speech Recognition Technology", IEEE, 2012.
5. Lindasalva Muda, "Voice Recognition Algorithm Using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", Journal of Computing, Vol. 2, Issue 3, March 2010.
6. Singh Satyanand, Dr. E. G. Rajan, "Vector Quantization Using MFCC and Inverted MFCC", International Journal of Computer Applications, Vol. 17, No. 1, pp. 1-7, March 2011.
7. Sun Xihao and Yoshikazi Miyanaga, "Dynamic time warping for speech recognition with training part to reduce the computation", IEEE, 2013.
8. Tudor Barbu, "Comparing various voice recognition techniques", IEEE, 2009.
9. Dawid Polap and Marcin Wozniak, "Voice Recognition by neuro-Heuristic Method", Tsinghua Science and Technology, Vol. 24, No. 1, February 2019.

## AUTHORS PROFILE



**Astha Gupta** is serving as an Assistant Professor in Chandigarh group of colleges, Landran and is having 2 years of teaching experience. She has done her Masters of Technology in Information technology from Punjab Technical University. She has done her research work in key areas of Visualization of Big data using Nanocubes. She has four research papers in international journals.



**Neeraj Sharma** is serving as Assistant Professor in Chandigarh group of colleges, Landran and is having 5 years of teaching experience. He has completed his Masters of Technology from Himachal Pradesh Technical University. He has done his research work in key areas of text mining using back propagation techniques. He has four research papers in international journals and have participated in two international conferences.

