

# **Speech Recognition with Advances in Artificial Intelligence**

**Vishakha Pawar and Priti Kudal**

Lecturer, Department of Computer Engineering,  
Guru Gobind Singh Polytechnic, Nashik, Maharashtra, India

**Abstract:** *This research study aims to present a surveying study about speech recognition systems and artificial intelligence. Artificial Intelligence is nothing but the demeanor by a machine completely as a human. Computer systems where AI is enabled can implement numerous errands just analogous to human or more often in efficient way. AI Systems are more time and energy saving systems. AI systems are proficient to give an error free response to convinced difficult situations. AI is transmuting the nature of nearly everything which is associated to human life e.g. communication, security, privacy, ethics, economic, healthcare etc. In immediate future, AI will carry out numerous activities such as, visual understanding, words identification, taking decisions, speech recognition, etc. Speech recognition has become one of the extensively used technologies, as it deals with great opportunity to intermingle and converse with automated machines. Specifically, it can be admitted that speech recognition enables its users and supports them to perform their daily routine tasks, in a more appropriate and effective manner. This research anticipates presenting the illustration of recent technological advancements, which are accompanying with artificial intelligence. Recent researches have exposed the fact that speech recognition is originate to be the supreme issue, which affects the cracking of speech. In direction to overcome these subjects, different statistical models were established by the researchers. Some of the most noticeable statistical models include acoustic model (AM), language model (LM), lexicon model, and hidden Markov models (HMM). The research will benefit in accepting all of these statistical models of speech recognition. Researchers have also expressed different decoding methods, which are being operated for realistic decoding tasks and controlled artificial languages. These decoding methods include pattern recognition, artificial intelligence, etc. It has been accepted that artificial intelligence is the most well-organized and trustworthy methods, which are being used in speech recognition.*

**Keywords:** Speech recognition, artificial intelligence, Hidden Markov Models (HMM), statistical models of speech recognition, human machine performance.

## **I. INTRODUCTION**

The main aim of this paper is to explain the various progression in artificial intelligence, in relation to speech recognition. It has been found from the analysis of research, which was given in [5] that speech recognition is the most ground-breaking concepts of electrical engineering and computer science. Primarily, this mode deals with the conversion of the spoken words into text. Speech recognition is abbreviated as STT (speech-to-text), ASR (automatic speech recognition), etc. But on the contrary, it has been said in [8] that speech recognition can also be understood as the field of computer science, which acts with the designing and development of computer systems, in direction to recognize the spoken words. In this regard, [14] has proclaimed that speech recognition or ASR is the method of renovating a speech signal into the sequence of words, by the support of various algorithms and procedures. It has been recognized in the research, which was carried out by [6] that these tactics comprise artificial intelligence approach, pattern recognition approach, and also acoustic phonetic approach. In accordance with the observations and opinions of [7], artificial intelligence is the most emerging and effective methods, which supports perfect and precise speech recognition. It is because; artificial intelligence integrates certain algorithmic methods, which promotes coherent

transformation and renovation of speech into decipherable patterns, and vice versa.

This research will support in accepting these concepts, which are allied with speech recognition. Among all of these approaches, artificial intelligence is established to be the most real and united approach, which has fortified and enhanced speech recognition practices [15]. The proceeding manuscript will creditably help in explaining the staple idea of artificial intelligence, as well as the hi-tech advances, which have been arisen in artificial intelligence. In addition to this, the paper will also contribute in understanding and recognizing the statistical models for speech recognition.

## **II. SPEECH RECOGNITION SYSTEMS**

Speech recognition can be anticipated as an approach, which pacts with the transformation of spoken words into the text. It has been recognized by [4] that speech recognition can also be denoted as ASR, as the method offers to identify the speech automatically. In agreement with the opinions and observations of [13], specific of the speech recognition systems applies speaker liberated speech recognition. On the other hand, news speech recognition systems uses training process, in which a speaker speaks segments of text into the speech recognition system. It has been recognized by [9] that these speech recognition systems identify the specific voice of a person and use it to change the recognition of person's speech; hence, resulting in extra intelligible and incorporated transcription. It is substantial to notice that schemes, which use training, and are mentioned as "speaker-dependent systems". However, systems, which do not use training, are known as "speaker-independent systems". It has been appealed by [12] that speech recognition can be smeared in various areas and sectors. Some of the most protuberant uses of speech recognition include speech-to-text processing, aircrafts, formulation of structured documents, smart search, domestic appliance control, call routing, simple data entry (credit card number entry), voice dialing, etc.

According to [11], incorporated and smooth presentation of speech recognition mostly is contingent on the exploitation of suitable statistical models. It is due to the information that these statistical models transform speech into decipherable form and vice versa [6]. Thereby, the acceptance and practice of unsuitable statistical model may disturb the reliability of speech recognition. Proceeding section includes the investigation of various statistical models, which are being used in speech recognition.

## **III. STATISTICAL MODELS OF SPEECH RECOGNITION**

### **A) Acoustic Model (AM)**

Acoustic model (AM) is most famous and used in speech recognition. It shows the characteristics of the basic recognition units. According to [14], the recognition units can be classified and used in 3 levels; phoneme level, syllable level, and at the word level. Many inadequacies and rules come into deliberation with the selection of each of the units. Reference [7] has shown that for LVCSR (large vocabulary continuous speech recognition) systems, phoneme is the most constructing and used unit. Hidden Markov models and neural networks (NN) are the adopted approaches in large scale, being used for the acoustic modeling of speech recognition systems.

### **B) Language Model (LM)**

Language model is next most noteworthy statistical model of speech recognition. One of the foremost objectives of language model is to express or transmit the behavior of the language. It is observed that it works with intends to broadcast the existence of the specific word sequences in the target speech. According to [6], the aspect of recognition engine, the language model, statistical model of speech recognition helps in minimizing the space required for investigation for a reliable and convincing combination of words. It is important to become aware of the language model that was developed in collaboration with CMU statistical LM toolkit.

### **C) Lexicon Model**

It has been observed by [10] that lexicon model uses the pronunciation of the words within the target speech, which has to be recognized. In accordance with the perceptions of [5], the lexicon model plays an important role in

ASR (Automatic Speech Recognition). Theory tells that the operations of lexical model are based on parameters, i.e., whole-word access, and decomposition of entire speech into tokens (small chunks). This process ultimately results in appropriate recognition of the speech. For instance, if the native language is used for speech recognition, the native language is used for formulation of lexicon model, in order to get valuable and useful results. In this regard, the greatest approach is artificial neural network phoneme, as it helps in developing the native lexicon from the foreign lexicon; hence, it helps in mapping the phone of English to the phones of native languages [12].

#### **D) Hidden Markov Models (HMM)**

Reference [14] has avowed that hidden Markov model is the most accepted statistical tool, which is used for the modeling and designing of data. It has been analyzed that hidden Markov model has played a vital role in minimizing the problems of speech classification, which was one of the core problems, within the speech recognition approach. According to [11], hidden Markov model integrated with various problems, which affects the accuracy of speech recognition. In order to solve those problems, weighted hidden Markov models and subspace projection algorithm were proposed [8]. Afterwards, this model has become the basis of modern HMM-based continuous speech recognition technology.

### **IV. DISTINCT DECODING METHODS OF SPEECH RECOGNITION**

It has been found in the research, which was carried out by [6] that number of techniques of decoding can be used, in order to recognize the speech. Acoustic phonetic method, pattern recognition method, and artificial intelligence approach [10] are the most outstanding and widely used methods. Above all methods are illustrated in the proceeding section.

#### **A) Pattern Recognition**

The widely adopted technique of speech recognition is pattern recognition. This method mostly incorporates two major steps, pattern comparison and pattern training. It has been recognized from the studies of [10] that the main characteristic of this method is that it uses a well-structured and integrated mathematical framework [5]. This mathematical framework helps in formulation of consistent representations of speech patterns; hence accurate results can be found. Pattern recognition has further two approaches, i.e., stochastic approach and template approach.

#### **B) Acoustic Phonetic**

According to [14], the most primary approaches of speech recognition were generally based on the routine of locating sounds and speeches. One of the major objectives of such activities was to provide enough labels to the sample sounds, in order to find the patterns of the sound. It is important to notice that such methods are found to be the foundation of the acoustic phonetic approach. As per the notion given by acoustic phonetic approach, there exist phonemes (phonetic units) and finite units within spoken language. These units of acoustic phonetic approach are the collection of acoustic properties that are mostly used as evidence in the speech signal.

#### **C) Artificial Intelligence**

According to [11], the approach of artificial intelligence is the one of the most used methods of speech recognition for decoding. Artificial intelligence can be understood as the pattern recognition approach in combination with acoustic phonetic approach. It is mainly due to the fact that it merges the concepts and ideas of pattern recognition methods and acoustic phonetic methods. It has been said that artificial intelligence is also known as knowledge based approach and it uses the information, which includes the spectrogram, phonetic, and linguistic. According to [1], the approach of artificial intelligence plays an important role in various activities of speech recognition, which includes demonstration of speech units, designing of recognition algorithm and representation of suitable appropriate inputs. It is important to notify that, among all methods of speech recognition, artificial intelligence is the most reliable and efficient methods.

Artificial intelligence can be defined as the emerging and continuously developing fields of computer science. It has been found that artificial intelligence mainly deals on the development of machines, which are capable to behave as human beings. It has been said in the research, carried out by [14] that these artificial machines responds in intelligent manner by collecting information from their respective environments, conniving appropriate and adequate steps, formulating answers, and presents required results. It has been said in the studies of [2]; [7], that artificial intelligence is mainly used in various areas, including traffic signals and lights, Smart homes equipments, maintenance systems and home security, healthcare, credit card, cell phones (smart phones), and video games. Besides these applications, artificial intelligence is expansively used in speech recognition.

#### **V. APPLICATION OF ARTIFICIAL INTELLIGENCE IN SPEECHRECOGNITION**

[3] presents that artificial intelligence is currently being worn in various distinct fields of life, including remote sensing, scientific discovery, transportation, law, aviation, robot control, medical diagnosis, etc. However, one of the exceptional applications of artificial intelligence is speech recognition. Studies of [8]; [12] show that the approach of artificial is mainly utilized in answering machines of call centre also customer care centre. In this account, [10] claims that speech recognition software make possible the computers to handle first level of NLP (Natural Language Processing), Data (Text) mining, and customer support, in order to enhanced and improved customer handling; which will result in customer satisfaction. Speech recognition find in the difficult issues, as it needs to have highly merged and considerate techniques. According to [12], in speech recognition, problems are often occurred due to the lack of vocabulary. In the current era, the methods of speech recognition has been used in various areas, including automated telephony systems, mobile phones, etc. However, the completion of error-free speech recognition, mainly for continuous speech, has remained an unsolved and major problem. Recent research proposed by [3], showed that different existing technologies have been developed by the researchers, which will attain rational accuracy of words. Specifically, emerging approaches and technological paradigms are having important role in slowly enhancing the speech recognition integrity.

On the divergent theme [7] has affirmed the fact that the accuracy of human listeners cannot compete with these technologies. Therefore, it is the most challenging errands for the researchers to design and develop unblemished and highly efficient speech recognition techniques. In these circumstances, the methods of artificial intelligence can be considered as most important and greatest opportunities, in terms of recognizing the patterns of speech. It is due to the fact that artificial intelligence incredibly transforms the speech into well- structured algorithms, by properly following all stages [15].

#### **VI. TECHNOLOGICAL ADVANCES IN THE FIELD OF ARTIFICIAL INTELLIGENCE**

Artificial intelligence has gained magnificent boost in the last two decades, as it has been practically used in extensive areas and fields of life. It has been exposed from the deep analysis of the research of [5] that the techniques of artificial intelligence have been advocated by many developers and researchers. Data mining, neural networks, fuzzy logic, and knowledge based systems are the most common and well-developed techniques of artificial intelligence. One of the important objectives behind this activity was to increase and improve the software development processes, in order to compete with today's fast paced environment. According to [13], some of the most given proof advancements in artificial intelligence include the development of graphical user interfaces, artificial neural network systems, object-oriented programming, natural language processing, rule based expert systems for aviation sector, fuzzy logic, speech recognition, text recognition, object recognition, robot navigation, intelligent transportation, as well as obstacle avoidance.

#### **VII. CONCLUSION**

From above discussion, it can be concluded that nowadays, the applications of speech recognition are flattering extensively vital role and are useful in all ways. It has found that speech recognition is the method of transforming the signals (input-Speech) into well-formed sequences of words, after conducting this research. It is important to notify that

these sequences are formed in the form of algorithms. Basically, these algorithms use to transform the speech to words and vice versa; hence results in more correct recognition of speech. It has been seen that speech recognition has become the greatest challenges, techniques and approaches that have been developed to overcome this problems. Along with all those paradigms and models, artificial intelligence is found to be most reliable and adequate methods. This research study has included the in-depth evaluation of the basic core concept of artificial intelligence. In addition to this, the paper has also illustrated the detailed speech recognition systems. Besides those different statistical models of speech recognition have also been explained in the paper, including acoustic model (AM), lexicon model, language model (LM) and Hidden Markov models (HMM).

These statistical models play a important role in designing the algorithms and patterns of speech. Additionally, it has also been found that different decoding methods are also used for speech recognition. Some of the most used methods include artificial intelligence, acoustic phonetic, and pattern recognition.

### REFERENCES

- [1]. Ammar, Hany H., WalidAbdelmoez, and Mohamed Salah Hamdi. "Software engineering using artificial intelligence techniques: Current state and open problems." Proceedings of the First Taibah University International Conference on Computing and Information Technology (ICCIT 2012), Al-Madinah Al-Munawwarah, Saudi Arabia. (2012), p.52, retrieved from, [http://www.researchgate.net/profile/Mohamed\\_Hamdi8/publication/254198356\\_Software\\_Engineering\\_Using\\_Artificial\\_Intelligence\\_Techniques\\_Current\\_State\\_and\\_Open\\_Problems/links/544771110cf2f14fb811f118.pdf](http://www.researchgate.net/profile/Mohamed_Hamdi8/publication/254198356_Software_Engineering_Using_Artificial_Intelligence_Techniques_Current_State_and_Open_Problems/links/544771110cf2f14fb811f118.pdf)
- [2]. Anusuya, A.M. and Katti, K.S. "Speech Recognition by Machine: A Review", (IJCSIS) International Journal of Computer Science and Information Security,(2009), p. 7, retrieved from, <http://arxiv.org/ftp/arxiv/papers/1001/1001.2267.pdf>
- [3]. Beigi, Homayoon. "Hidden Markov Modeling (HMM)." Fundamentals of Speaker Recognition. Springer,(2011), p.41,retrieved from, [http://link.springer.com/chapter/10.1007/978-0-387-77592-0\\_13](http://link.springer.com/chapter/10.1007/978-0-387-77592-0_13)
- [4]. Besacier, Laurent, et al. "Automatic speech recognition for under- resourced languages: A survey." Speech Communication, (2014), p. 100, retrieved from, <http://www.sciencedirect.com/science/article/pii/S0167639313000988>
- [5]. Chen, Chi-hau, ed. "Pattern recognition and artificial intelligence." Elsevier, (2013), p. 6, retrieved from, [https://books.google.com/books?hl=en&lr=&id=QixXL8sxJgEC&oi=fnd&pg=PP1&dq=SPEECH+RECOGNITION+SYSTEM+and+artificial+intelligence&ots=Z7cz1bl-YF&sig=9qQWy2oto\\_ijLb9Dig1Onby7M5E&v=onepage&q=SPEECH%20RECOGNITION%20SYSTEM%20and%20artificial%20intelligence&f=false](https://books.google.com/books?hl=en&lr=&id=QixXL8sxJgEC&oi=fnd&pg=PP1&dq=SPEECH+RECOGNITION+SYSTEM+and+artificial+intelligence&ots=Z7cz1bl-YF&sig=9qQWy2oto_ijLb9Dig1Onby7M5E&v=onepage&q=SPEECH%20RECOGNITION%20SYSTEM%20and%20artificial%20intelligence&f=false)
- [6]. Chen, Lijiang, et al. "Speech emotion recognition: Features and classification models." Digital signal processing 22.6 (2012), p. 15, retrieved from, <http://www.sciencedirect.com/science/article/pii/S1051200412001133>
- [7]. Choudhary, A. and Kshirsagar, R. "Process Speech Recognition System using Artificial Intelligence Technique, International Journal of Soft Computing and Engineering (IJSCE), (2012), p. 3,retrievedfrom, <http://www.ijsee.org/attachments/File/v2i5/E1054102512.pdf>
- [8]. Dalby, Jonathan, and Diane Kewley-Port. "Explicit pronunciation training using automatic speech recognition technology." Calico Journal, (2013), p. 22, retrieved from,<http://www.equinoxpub.com/journals/index.php/CALICO/article/viewArticle/23361>
- [9]. Deng, Li, and Xiao Li. "Machine learning paradigms for speech recognition: An overview." IEEE Transactions on Audio, Speech and Language Processing 21.5 (2013), p. 45, retrieved from, [http://131.107.65.14/pubs/189008/tasl-deng-2244083-x\\_2.pdf](http://131.107.65.14/pubs/189008/tasl-deng-2244083-x_2.pdf)
- [10]. Hinton, Geoffrey, et al. "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups." Signal Processing Magazine, IEEE 29.6 (2012), pp. 82-97, retrieved from, <http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6296526&url=ht tp%3A%2F%2Fieeexplore.ieee.org>

- %2Fxppls%2Fabs\_all.jsp%3Farnumber%3D6296526
- [11]. Mikolov, Tomas. "Statistical language models based on neural networks." Presentation at Google, Mountain View, (2012), p. 7, retrieved from, <http://www.fit.vutbr.cz/~imikolov/rnnlm/google.pdf>
  - [12]. Morgan, Nelson. "Deep and wide: Multiple layers in automatic speech recognition." Audio, Speech, and Language Processing, IEEE Transactions on 20.1 (2012), p. 6, retrieved from, [http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5714717&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs\\_all.jsp%3Farnumber%3D5714717](http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5714717&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D5714717)
  - [13]. Rawat, Seema, Parv Gupta, and Praveen Kumar. "Digital life assistant using automated speech recognition." IEEE, (2014), p. 13, retrieved from [http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=7019075&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs\\_all.jsp%3Farnumber%3D7019075](http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=7019075&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D7019075).
  - [14]. Saini, Preeti, and Parneet Kaur. "Automatic Speech Recognition: A Review." International journal of Engineering Trends & Technology (2013), pp. 132-136, retrieved from, <http://www.ijettjournal.org/volume-4/issue-2/IJETT-V4I2P210.pdf>
  - [15]. Saon, George, and Jen-TzungChien. "Large-vocabulary continuous speech recognition systems: A look at some recent advances." Signal Processing Magazine, IEEE 29.6 (2012), p. 18, retrieved from, [http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6296522&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs\\_all.jsp%3Farnumber%3D6296522](http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6296522&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D6296522).