

Artificial Intelligence Advances Through Speech Recognition

Dr. Rajendra Kumar Mahto

Assistant Professor

Department Of Information Technology,
Dr Shayama Prasad Mukherjee University, Ranchi

rajendrabit57@gmail.com

Abstract

The purpose of this research study is to provide a retrospective analysis of artificial intelligence and speech recognition systems.

Since speech recognition provides a great opportunity to interact and communicate with automated machines, it has become one of the most widely used technologies. It can be stated with precision that speech recognition makes life easier for its users and enables them to carry out their regular activities in a more efficient and convenient way. The purpose of this study is to provide an example of current technological developments related to artificial intelligence.

Recent studies have demonstrated that speech recognition is the most critical factor affecting speech decoding. The researchers created various statistical models in order to get around these problems. A selection of the most well-known statistical models comprise hidden Markov models (HMM), lexicon models, language models (LM), and acoustic models (AM). Understanding each of these statistical models of speech recognition will be aided by the research.

Additionally, researchers have developed various decoding techniques that are applied to constrained artificial languages and realistic decoding tasks. These decoding techniques include artificial intelligence, acoustic phonetics, and pattern recognition. Artificial intelligence has been acknowledged as the most dependable and efficient technique for speech recognition.

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

I. INTRODUCTION

The aim of this paper is to demonstrate various advancements in artificial intelligence from the standpoint of speech recognition. It possesses the analysis of research conducted by [5] has demonstrated that speech recognition represents one of the most sophisticated ideas in computer science and electrical engineering. This method essentially focuses on turning spoken words into written language. Other names for speech recognition include computer speech recognition, speech to text (STT), and automatic speech recognition (ASR). As an alternative, [8] asserts that speech recognition is also a branch of computer science that studies the creation and design of computer systems for the purpose of talking word recognition. According to [14], speech recognition, also known as computer speech recognition or ASR, is just the process of converting a Speech signal is converted into a word sequence using a variety of methods and algorithms. These approaches include the artificial intelligence approach, the pattern recognition approach, and the acoustic phonetic approach, according to research conducted by [6]. As per the opinions and understandings of [7], artificial intelligence is one of the most advanced and efficient methods that facilitates precise and faultless speech recognition. The reason for this is that artificial intelligence uses specific algorithmic techniques that support the coherent conversion and transformation of speech into readable patterns, and vice versa.

These ideas, which are connected to speech recognition, will be better understood with the help of this research. The most successful and comprehensive strategy among all of these is artificial intelligence, which has enhanced and reinforced speech recognition procedures [15]. The manuscript that will be presented will be of

great assistance in explaining the fundamental idea of artificial intelligence and the technological advancements that have been made in this field.

Furthermore, the paper will aid in the comprehension and identification of statistical models related to speech recognition.

II. SPEECH RECOGNITION SYSTEM

One way to think of speech recognition is as a method that translates spoken words into written language. According to [4], speech recognition is also known as automatic speech recognition (ASR) because this technology allows speech to be recognized automatically. Some speech recognition systems make use of speaker independent speech recognition, in line with the opinions and perceptions of [13]. Conversely, however, certain speech recognition systems employ a training technique wherein a human speaker reads text passages to the speech recognition system. According to [9], these speech recognition systems are able to identify an individual's unique voice and use that information to adjust the speech recognition, leading to transcriptions that are more cohesive and well-rounded. It is noteworthy to observe that systems that rely on training are called "speaker-dependent systems." On the other hand, these systems—which do not require training—are referred to as "speaker-independent systems." According to [12], speech recognition has applications in a variety of fields and industries. The most well-known uses of speech recognition technology are in aircraft (direct voice input), speech-to-text processing (emails or work processors), and the creation of structured documents (such as radiology reports), straightforward data entry (such as entering a credit card number), intelligent search (podcast), control over household appliances, call routing, voice dialling, etc.

As per the findings of [11], seamless and cohesive execution of Advances in Artificial Intelligence Through Speech Recognition.

Speech recognition is primarily dependent on the application of suitable statistical models. The reason for this is that these statistical models translate speech into readable form and the other way around [6]. Therefore, the integrity of speech recognition may be impacted by the adoption and application of an inappropriate statistical model. The examination of various statistical models used in speech recognition is included in the section that follows.

III. STATISTICAL MODELS OF SPEECH RECOGNITION

A. Model of Acoustics (AM)

The acoustic model (AM) is one of the most well-known and extensively used speech recognition models. It has been shown that the features of the fundamental recognition units are captured by acoustic speech recognition models. As per reference [14], recognition units can be identified at the phoneme, syllable, and word levels. The choice of each of these units takes into account a number of shortcomings and limitations. According to reference [7], the phoneme is the most advantageous unit for LVCSR (large vocabulary continuous speech recognition) systems. Neural networks (NN) and hidden Markov models are the two most popular methods for modelling speech recognition systems' acoustic characteristics.

B. Model of Language (LM)

One of the most important statistical models for speech recognition is the language model. One of the main goals of the language model is to transmit or convey the behaviour of the tongue. This is because its goal is to predict whether or not the target speech will contain the particular word sequences. [6] states that this statistical model of speech recognition helps to reduce the search space for a trustworthy and credible word combination from the perspective of the recognition engine. It is important to note that the CMU statistical LM toolkit was used in the development of the language model.

C. Lexicon Model

According to [10], the lexicon model gives the words in the target speech their pronunciation, which needs to be identified. According to the opinions expressed in [5], the lexicon model is essential to automatic speech recognition. It results from the fact that the reason behind this is that the lexical model's operations rely on two parameters: whole-word access and the breaking down of speech into smaller segments. The speech is eventually recognized appropriately as a result of this process. For example, in order to obtain valuable and

practical results, the lexicon model must be formulated in the native languages if speech recognition models are in those languages. Artificial neural network phoneme is regarded as one of the best methods in this sense since it helps create the native vocabulary from the foreign lexicon, which leads to the mapping of the English phoneme to the native language phonemes [12]. It is crucial to observe that the entire procedure is carried out while taking the contextual

D. Information about Hidden Markov Models (HMM)

Reference [14] has confirmed that the HMM is the most widely used statistical tool for data modeling. Analysis has shown that the hidden Markov model has significantly contributed to the resolution of speech classification problems, which were one of the main problems with the speech recognition approach. [11] states that the hidden Markov model took into account a number of problems that formerly affected the speech recognition accuracy. Weighted hidden Markov models and the subspace projection algorithm were proposed as solutions to those problems [8]. Subsequently, this model served as the foundation for contemporary continuous speech recognition technology based on HMMs.

IV. DECENTRATIVE DECODING TECHNIQUES FOR SPEECH RECOGNITION

It has been confirmed by the studies, which was conducted by [6] revealed that a variety of decoding methods could be applied to identify the speech. The artificial intelligence approach, pattern recognition method, and acoustic phonetic method are a few of the most well-known and frequently applied techniques [10]. The section that follows provides a brief illustration of each of these techniques.

A. Identification of Patterns

One of the most popular and extensively used methods for speech recognition is pattern recognition. Two crucial steps are included in this method: pattern training and pattern comparison. Based on research conducted by [10], it has been determined that the primary feature of this approach is its utilization of a well-organized and cohesive mathematical framework [5]. In order to create consistent representations of speech patterns, this mathematical framework is helpful. lead to the acquisition of more precise outcomes. Two additional approaches to pattern recognition are the stochastic approach and the template approach.

B. Phonetic Acoustic

As per reference [14], the earliest methods of speech recognition relied primarily on the identification of sounds and speech patterns. Providing appropriate labels for the sample sounds was one of the main goals of these activities, as it allowed participants to identify the sound patterns. It's important to note that these techniques are what the acoustic phonetic approach is built upon. According to the concept of the acoustic phonetic approach, spoken language consists of phonemes, or phonetic units, and finite units. The vast majority of these acoustic phonetic approach units are classified by the assortment of acoustic attributes that are typically audible in the speech signal.

C. Artificial Intelligence

As per reference [11], the artificial intelligence approach is one of the most well-known speech recognition techniques utilized for decoding. It is possible to think of artificial intelligence as the result of combining the acoustic phonetic approach and pattern recognition approach. The reason for this is that the International Journal of Computer and Information Engineering, Vol. 9, No. 6, 2015, 1440, World Academy of Science, Engineering, and Technology Worldwide Academic and Scientific Investigation & Development 9(6) 2015scholar.waset.org/1307-6892/10001552

The International Science Index, Computer and Information Engineering Vol:9, No:6, 2015 waset.org/Publication/10001552 states that it integrates the ideas and concepts of acoustic phonetic approach and pattern recognition techniques. It has been demonstrated that artificial intelligence, also known as the knowledge-based approach, makes use of the data, which has spectrogram, phonetic, and linguistic connections.

As per reference [1], the artificial intelligence approach is deemed indispensable in various speech recognition activities such as algorithm design, speech unit demonstration, and representation of appropriate and suitable inputs. It is important to note that artificial intelligence is the most reliable and effective method of speech recognition among all of the methods.

Artificial intelligence can be broadly defined as the rapidly developing fields of computer science that are still in their infancy. According to analysis, the main focus of artificial intelligence is on creating machines that can mimic human behaviours.

Research on the subject has been documented, with work done by [14] stating that these artificial machines gather data from their surroundings and react intelligently, figuring out the right course of action, coming up with solutions, and displaying the intended outcomes. Based on research studies [2]; [7], it has been established that artificial intelligence is extensively utilized in various domains, such as traffic lights and pedestrian signals, robotic home appliances, home security and maintenance systems, credit card transactions, healthcare robotics, mobile phones (smart phones), and video games. In addition to these uses, speech recognition is one area where artificial intelligence is heavily utilized.

V. APPLICATION OF ARTIFICIAL INTELLIGENCE IN SPEECH RECOGNITION

Based on the analysis of research presented by [3], it has been noted that artificial intelligence is currently being used in a variety of fields, such as scientific discovery, remote sensing, aviation, toys, law, stock trading, transportation, robot control, and even medical diagnosis. Speech recognition, however, is one of the most remarkable uses of artificial intelligence. Research from [8] and [12] demonstrates that the artificial intelligence method is widely used in call centers and customer service answering machines. According to this account, [10] asserts that speech recognition software empowers computers to perform basic natural language processing, text mining, and customer support, thereby promoting enhanced and improved customer handling and ultimately increasing customer satisfaction.

One of the more challenging problems is speech recognition, which requires extremely thoughtful and integrated techniques.

As per reference [12], difficulties with speech recognition are frequently caused by insufficient vocabulary. In the present speech recognition technology has been applied in a variety of contexts in the modern era, such as automated telephone systems, mobile phones, etc. Nevertheless, the achievement of error-free speech recognition has proven to be a challenging problem, particularly for continuous speech. According to recent research, which was put forth by [3], researchers have developed a variety of modern technologies that have enabled them to achieve a reasonable level of word accuracy.

More specifically, new methodologies and technological paradigms are making a noteworthy contribution to the consistent improvement of speech recognition integrity. However, it has been stated by [7] that these technologies are insufficiently powerful to match the accuracy of human listeners.

For this reason, designing and developing is one of the hardest things for researchers to do. perfect and incredibly effective speech recognition systems. In these kind of situations, one of the biggest opportunities for accurately identifying speech patterns is the application of artificial intelligence. The reason for this is because artificial intelligence effectively converts speech into well-structured algorithms by suitably adhering to each step [15]. The most crucial phases of speech recognition using artificial intelligence are speech unit representation, recognition algorithm formulation and development, and audio demonstration of accurate inputs.

VI. TECHNOLOGICAL DEVELOPMENTS IN THE ARTIFICIAL INTELLIGENCE FIELD

Over the past 20 years, artificial intelligence has advanced significantly due to its widespread application in many spheres of life. It has been made clear by the thorough examination of the research of [5] reveals that a number of developers and scientists have supported artificial intelligence techniques. Neural networks, fuzzy logic, data mining, and knowledge-based systems are a few of the most popular and advanced artificial intelligence techniques.

Enhancing and improving the software development processes was one of the main goals of this activity, as it is necessary to compete in the fast-paced and unpredictable environment of today. The development of artificial neural network systems, graphical user interfaces, object-oriented programming, natural language processing, fuzzy logic, rule-based expert systems for the aviation sector, speech recognition, text recognition, robot

navigation, object recognition, intelligent transportation, and obstacle avoidance are some of the most obvious advancements in artificial intelligence, according to [13].

VII. CONCLUSION

Based on the preceding discourse, it can be inferred that the uses of speech recognition technology are increasingly significant and practical in today's world. It has been determined through the course of this research that speech recognition is the process of converting input signals—typically speech—into coherent word sequences. The development of these sequences in the form of algorithms is an important observation. These algorithms, in essence, translate speech into words and vice versa, leading to more accurate, consistent, and coherent speech recognition. According to assessments, one of the biggest obstacles nowadays is speech recognition, and a number of methods and strategies have been created in order to get around this problem. Among all of those models and paradigms, artificial intelligence is regarded as one of the most suitable and dependable strategies. This research study has integrated a thorough analysis of the fundamental idea of artificial intelligence. Furthermore, the paper has provided illustrations of the speech recognition systems.

In addition, several statistical models of speech recognition, such as the lexicon model, acoustic model (AM), language model (LM), and Hidden Markov models (HMM), have also been summarized in the paper.

These statistical models are crucial in the development of the speech patterns and algorithms that must be identified.

Furthermore, it has additionally been found that various decoding techniques are also applied to speech recognition. Artificial intelligence, acoustic phonetics, and pattern recognition are a few of the most popular and widely used techniques. Out of all these strategies or tactics, artificial intelligence can be regarded as the most well-rounded and efficient strategy since it yields incredibly accurate and dependable results. Along with evaluating the technological advancements in the field of artificial intelligence, the paper also shows how artificial intelligence is applied in speech recognition.

REFERENCES:

- [1] Mohamed Salah Hamdi, Walid Abdelmoez, Mohamed Ammar, and Hany H. "Software engineering using artificial intelligence techniques: Current state and open problems." Record of the First International Conference on Computing and Information Technology (ICCIT 2012), Taibah University, Al-Madinah Al-Munawwarah, Saudi Arabia, p. 52, extracted from http://www.researchgate.net/profile/Mohamed_Hamdi8/publication/254198356_Software_Engineering_Using_Artificial_Intelligence_Techniques_Current_State_and_Open_Problems/links/544771110cf2f14fb811f118.pdf
- [2] Speech Recognition by Machine: A Review, Anusuya, A.M. and Katti, K.S. International Journal of Computer Science and Information Security, (IJCSIS) (2009), p. 7, retrieved from <http://arxiv.org/ftp/arxiv/papers/1001/1001.2267.pdf>
- [3] Homayoon, Beigi. "Hidden Markov Modeling (HMM)." The Essentials of Speech Recognition. Retrieved from http://link.springer.com/chapter/10.1007/978-0-387-77592-0_13 by Springer (2011), page 41
- [4] "Automatic speech recognition for under-resourced languages: A survey," by Laurent Besacier et al. S0167639313000988, Speech Communication, (2014), p. 100, was obtained from <http://www.sciencedirect.com/science/article/pii/S0167639313000988>
- [5] Chen, Chi-hau, ed. Artificial intelligence and pattern recognition. Elsevier, (2013), p. 6, obtained from <https://books.google.com/books?>
- [6] Chen, Lijiang, et al. "Features and classification models for speech emotion recognition." retrieved from <http://www.sciencedirect.com/science/article/pii/S1051200412001133>, Digital Signal Processing 22.6 (2012), p.

- [7] Kshirsagar, R. and Choudhary, A. "Process International Journal of Soft Computing and Engineering (IJSCE), (2012), p. 3, Speech Recognition System utilizing Artificial Intelligence Technique, retrieved from <http://www.ijsce.org/attachments/File/v2i5/E1054102512.pdf>
- [8] Jonathan Dalby and Diane Kewley-Port. "Explicit pronunciation training using automatic speech recognition technology." Retrieved from <http://www.equinoxpub.com/journals/index.php/CALICO/article/viewArticle/23361>, Calico Journal, 2013, p. 22
- [9] Deng, Li, and Xiao Li. "An Overview of Machine Learning Paradigms for Speech Recognition." IEEE Transactions on Audio, Speech, and Language Processing 21.5 (2013), retrieved from http://131.107.65.14/pubs/189008/tasl-deng-2244083-x_2.pdf, page 45
- [10] Geoffrey Hinton et al. "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups." 82–97, IEEE 29.6 (2012), Signal Processing Magazine, retrieved from http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6296526&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D6296526
- [11] The publication "Statistical language models based on neural networks" was written by Mikolov, Tomas. retrieved from the 2012 Mountain View, CA presentation at Google, page 7. via, Google.pdf at <http://www.fit.vutbr.cz/~imikolov/rnnlm>.
- [12] Morgan, Nelson. "Many layers in automatic speech recognition: Deep and wide." Audio, Speech, and Language Processing, IEEE Transactions on 20.1 (2012), p. 6; obtained from http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5714717&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D5714717
- [13] Praveen Kumar, Seema Rawat, Parv Gupta, and Parv Gupta. "Automated speech recognition for digital life assistant." IEEE, 2014, p. 13. downloaded from 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. "Automated Speech Recognition: A Comprehensive Analysis." 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-Tzung Chien. "A review of some recent developments in large-vocabulary continuous speech recognition systems." IEEE 29.6 (2012), p. 18, Signal Processing Magazine, 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.