World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:9, No:6, 2015

Advances in Artificial Intelligence Using Speech Recognition

Khaled M. Alhawiti

Open Science Index, Computer and Information Engineering Vol:9, No:6, 2015 publications.waset.org/10001552/pdf

***Abstract*—**This research study aims to present a retrospective study about speech recognition systems and artificial intelligence. Speech recognition has become one of the widely used technologies, as it offers great opportunity to interact and communicate with automated machines. Precisely, it can be affirmed that speech recognition facilitates its users and helps them to perform their daily routine tasks, in a more convenient and effective manner. This research intends to present the illustration of recent technological advancements, which are associated with artificial intelligence. Recent researches have revealed the fact that speech recognition is found to be the utmost issue, which affects the decoding of speech. In order to overcome these issues, different statistical models were developed by the researchers. Some of the most prominent statistical models include acoustic model (AM), language model (LM), lexicon model, and hidden Markov models (HMM). The research will help in understanding all of these statistical models of speech recognition. Researchers have also formulated different decoding methods, which are being utilized for realistic decoding tasks and constrained artificial languages. These decoding methods include pattern recognition, acoustic phonetic, and artificial intelligence. It has been recognized that artificial intelligence is the most efficient and reliable methods, which are being used in speech recognition.

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

I. Introduction

THE purpose of this paper is to present the illustration of different advancement in artificial intelligence, in the perspective of speech recognition. It has been established from the analysis of research, which was conducted by [5] that speech recognition is one of the most advanced concepts of electrical engineering and computer science. Basically, this approach deals with the conversion of the spoken words into text. Speech recognition is also referred as ASR (automatic speech recognition), STT (speech to text) or just computer speech recognition. On the contrary, it has been claimed by [8] that speech recognition can also be understood as the field of computer science, which deals with the designing and development of computer systems, in order to recognize the spoken words. In this regard, [14] has asserted that speech recognition or computer speech recognition or ASR is nothing more than the approach of converting a speech signal into the sequence of words, by the help of different algorithms and techniques. It has been documented in the research, which was carried out by [6] that these approaches include artificial intelligence approach, pattern recognition approach, as well as acoustic phonetic approach. In accordance with the views and perceptions of [7], artificial intelligence is the most developing and effective techniques, which supports flawless and accurate speech recognition. It is because; artificial intelligence incorporates certain algorithmic approaches, which fosters coherent conversion and transformation of speech into readable patterns, and vice versa.

This research will assist in understanding these concepts, which are associated with speech recognition. Amid all of these approaches, artificial intelligence is found to be the most effective and integrated approach, which has strengthened and improved speech recognition practices [15]. The proceeding manuscript will commendably help in illustrating the core concept of artificial intelligence, as well as the technological advances, which have been occurred in artificial intelligence. In addition to this, the paper will also assist in understanding and identifying the statistical models for speech recognition.

Khaled M. Alhawiti is with the Tabuk University, Tabuk, Saudi Arabia (phone: +9660555903348; e-mail: alhowity@hotmail.com).

II. Speech Recognition Systems

Speech recognition can be understood as an approach, which deals with the translation of spoken words into the text. It has been established by [4] that speech recognition can also be referred as ASR, as the technique offers to recognize the speech automatically. In accordance with the views and perceptions of [13], some of the speech recognition systems utilize speaker independent speech recognition. On the other hand, other speech recognition systems utilize training method, in which an indivuial speaker reads sections of text into the speech recognition system. It has been established by [9] that these speech recognition systems recognize the specific voice of a person and use it to modify the recognition of person's speech; hence, resulting in more coherent and integrated transcription. It is significant to notice that systems, which use training, are referred as "speaker-dependent systems". However, such systems, which do not use training, are known as "speaker-independent systems". It has been claimed by [12] that speech recognition can be applied in different areas and sectors. Some of the most prominent applications of speech recognition include aircrafts (direct voice input), speech-to-text processing (emails or work processors), formulation of structured documents (radiology reports), simple data entry (credit card number entry), smart search (podcast), domestic appliance control, call routing, voice dialing, etc.

According to [11], integrated and smooth performance of

Open Science Index, Computer and Information Engineering Vol:9, No:6, 2015 publications.waset.org/10001552/pdf

speech recognition mainly depends on the utilization of appropriate statistical models. It is due to the fact that these statistical models convert speech into readable form and vice versa [6]. Thereby, the adoption and usage of inappropriate statistical model may affect the integrity of speech recognition. Proceeding section incorporates the analysis of different statistical models, which are being used in speech recognition.

III. Statistical Models of Speech Recognition

A. Acoustic Model (AM)

One of the most prominent and widely adopted models of speech recognition is acoustic model (AM). It has been established that acoustic models of speech recognition capture the characteristics of the basic recognition units. According to [14], the recognition units can be at the phoneme level, syllable level, and at the word level. Several inadequacies and constraints come into consideration with the selection of each of these units. Reference [7] has claimed that for LVCSR (large vocabulary continuous speech recognition) systems, phoneme is the most favorable unit. Hidden Markov models and neural networks (NN) are the widely adopted approaches, which are being utilized for the acoustic modeling of speech recognition systems.

B. Language Model (LM)

Language model is another most significant statistical model of speech recognition. One of the major objectives of language model is to convey or transmit the behavior of the language. It is due to the fact that it intends to forecast the existence of the specific word sequences within the target speech. According to [6], from the aspect of recognition engine, this statistical model of speech recognition assists in minimizing the search space for a reliable and credible combination of words. It is significant to notice that language model was developed by the help of CMU statistical LM toolkit.

C. Lexicon Model

It has been claimed by [10] that lexicon model provides 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 inevitable and indispensable role in automatic speech recognition. It is due to the fact that the operations of lexical model are based on two parameters, i.e., whole-word access, and decomposition of entire speech into small chunks. This process eventually results in appropriate recognition of the speech. For instance, if speech recognition models are in native language, the lexicon model has to be formulated in the native languages, in order to acquire valuable and useful results. In this regard, artificial neutral network phoneme can be considered as one of the greatest approaches, as it assists in developing the native lexicon from the foreign lexicon; hence, resulting in mapping the phone of English to the phones of native languages [12]. It is important to notice that the entire process is conducted, while considering the contextual information.

D. Hidden Markov Models (HMM)

Reference [14] has affirmed that hidden Markov model is the most popular statistical tool, which is being used for the modeling of data. It has been analyzed that hidden Markov model has played a commendable role in reducing the issues of speech classification, which was one of the core issues, within the speech recognition approach. According to [11], hidden Markov model incorporated various issues, which used to affect the accuracy of speech recognition. In order to resolve those issues, subspace projection algorithm and weighted hidden Markov models were proposed [8]. Afterwards, this model has become the basis of modern HMM-based continuous speech recognition technology.

IV. Distinctive Decoding Methods of Speech Recognition

It has been documented in the research, which was carried out by [6] that various different techniques of decoding can be used, in order to recognize the speech. Some of the most prominent and widely used methods include acoustic phonetic method, pattern recognition method, and artificial intelligence approach [10]. All of these methods are briefly illustrated in the proceeding section.

A. Pattern Recognition

Pattern recognition is found to be the most common and widely adopted techniques of speech recognition. This method mainly incorporates two important steps, including pattern comparison and pattern training. It has been established from the studies of [10] that the leading characteristic of this method is that it utilizes a well structured and integrated mathematical framework [5]. This mathematical framework assists in formulating consistent representations of speech patterns; hence result in the acquisition of more accurate results. Pattern recognition is further divided into two more approaches, i.e., stochastic approach and template approach.

B. Acoustic Phonetic

According to [14], the most primitive approaches of speech recognition were mainly based on the process of locating sounds and speeches. One of the major objectives of such activities was to provide adequate labels to the sample sounds, in order to recognize the patterns of the sound. It is significant to notice that such methods are found to be the foundation of the acoustic phonetic approach. As per the notion of acoustic phonetic approach, there exist phonemes (phonetic units) and finite units within spoken language. These units of acoustic phonetic approach are extensively categorized by the collection of acoustic properties that are usually evident in the speech signal.

C. Artificial Intelligence

According to [11], the approach of artificial intelligence is the most famous methods of speech recognition, which are being used for decoding. Artificial intelligence can be understood as the combination of the pattern recognition approach and acoustic phonetic approach. It is due to the fact

Open Science Index, Computer and Information Engineering Vol:9, No:6, 2015 publications.waset.org/10001552/pdf

that it incorporates the concepts and ideas of pattern recognition methods and acoustic phonetic approach. It has been established that artificial intelligence is also referred as knowledge based approach and it utilizes the information, which is related to spectrogram, phonetic, and linguistic. According to [1], the approach of artificial intelligence plays an indispensable role in different activities of speech recognition, including designing of recognition algorithm, demonstration of speech units, and representation of suitable and appropriate inputs. It is significant to bring into the notice that, among all methods of speech recognition, artificial intelligence is the most credible and efficient methods.

In general, artificial intelligence can be understood as the emerging and continually developing fields of computer science. It has been analyzed that artificial intelligence mainly emphasizes on the development of such machines, which are capable enough to get engage in the behaviors of human beings.

It has been documented in the research, which was carried out by [14] that these artificial machines collects information from their respective environments and respond in an intelligent manner, calculating appropriate and adequate steps, formulating answers, and presents desired results. It has been recognized from the studies of [2]; [7], that artificial intelligence is widely used in different areas, including pedestrian signals and traffic lights, robotic household equipment, maintenance systems and home security, healthcare robotics, credit card transactions, cell phones (smart phones), and video games. Besides all of these applications, artificial intelligence is extensively used in speech recognition.

V. Application of Artificial Intelligence in Speech Recognition

It has been observed from the evaluation of studies, which presented by [3] that artificial intelligence is currently being used in different fields of life, including scientific discovery, remote sensing, transportation, aviation, law, robot control, stock trading, medical diagnosis, and even toys. However, one of the most outstanding applications of artificial intelligence is speech recognition. Studies of [8]; [12] show that the approach of artificial is broadly utilized in answering machines of customer care centres and call centres. In this account, [10] claims that speech recognition software enable the computers to handle first level of natural language processing, text mining, and customer support, in order to foster enhanced and improved customer handling; hence results in customer satisfaction.

Speech recognition is one of the difficult issues, as it needs to have highly integrated and considerate techniques. According to [12], in speech recognition, issues are often occurred due to the lack of ample vocabulary. In the current era, the approach of speech recognition has been used in different areas, including automated telephony systems, mobile phones, etc. However, the accomplishment of error-free speech recognition, specifically for continuous speech, has remained an unsolved and difficult issue. Recent research, which was proposed by [3], shows that different contemporary technologies have been developed by the researchers, which have made it possible to attain reasonable accuracy of words. Precisely, emerging approaches and technological paradigms are playing a commendable role in steadily enhancing the integrity of speech recognition. On the contrary [7] has declared the fact that these technologies are not capable enough to compete with the accuracy of human listeners. Therefore, it is one of the most challenging tasks for the researchers to design and develop flawless and highly efficient speech recognition techniques. In such circumstances, the approach of artificial intelligence can be considered as one of the greatest opportunities, in terms of recognizing the patterns of speech, accurately. It is due to the fact that artificial intelligence incredibly transforms the speech into well-structured algorithms, by appropriately following all stages [15]. Most important stages, which are involved in speech recognition through artificial intelligence includes representation of speech units, formulation and development of recognition algorithms, as well as demonstration of correct inputs (speech).

VI. Technological Advances in the Field of Artificial Intelligence

Artificial intelligence has gained tremendous boost in the past two decades, as it has been applied in extensive areas and fields of life. It has been revealed from the profound analysis of the research of [5] that the techniques of artificial intelligence have been advocated by several developers and researchers. Some of the most common and well-developed techniques of artificial intelligence includes data mining, fuzzy logic, neural networks, and knowledge based systems. One of the major objectives behind this activity was to enhance and improve the software development processes, in order to compete with today's fast paced and volatile environment. According to [13], some of the most evident advancements in artificial intelligence include the development of artificial neural network systems, graphical user interfaces, object-oriented programming, natural language processing, fuzzy logic, rule based expert systems for aviation sector, speech recognition, text recognition, robot navigation, object recognition, intelligent transportation, as well as obstacle avoidance.

VII. Conclusion

From above discussion, it can be concluded that the applications of speech recognition are becoming extensively important and useful nowadays. After conducting this research, it has been established that speech recognition is the process of transforming the input signals (usually speech) into the well-structured sequences of words. It is significant to notice that these sequences are developed in the form of algorithms. Basically, these algorithms converts the speech into the words and vice versa; hence results in more coherent, accrete, and correct recognition of speech. It has been assessed that speech recognition has become one of the greatest challenges and several techniques and approaches have been

developed, to overcome this issue. Amid all of those paradigms and models, artificial intelligence is considered as one of the most reliable and adequate approaches. This research study has incorporated the in-depth evaluation of the core concept of artificial intelligence. In addition to this, the paper has also illustrated the speech recognition systems. Besides that different statistical models of speech recognition have also been encapsulated in the paper, including acoustic model (AM), language model (LM), lexicon model, and Hidden Markov models (HMM).

These statistical models play a major role in designing the algorithms and patterns of speech, which has to be recognized. Additionally, it has also been analyzed that different decoding methods are also used for speech recognition. Some of the most common and extensively adopted methods include artificial intelligence, acoustic phonetic, and pattern recognition. Amid all of these approaches or methods, artificial intelligence can be considered as the most integrated and effective approaches, as this technique provides highly reliable and accurate results. The paper has also demonstrated the application of artificial intelligence in speech recognition, while assessing the technological advances in the field of artificial intelligence.

Open Science Index, Computer and Information Engineering Vol:9, No:6, 2015 publications.waset.org/10001552/pdf

REFERENCES

[1] Ammar, Hany H., Walid Abdelmoez, 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/5447711l0cf2f14fb811f118.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

[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

[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, retrieved from, http://www.ijsce.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 Processing21.5 (2013), p. 45, retrieved from, 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=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%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 on20.1 (2012), p. 6, retrieved from, http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=5714717&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%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%2Fxpls%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-Tzung Chien. "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%2Fxpls%2Fabs_all.jsp%3Farnumber%3D6296522.