ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 9, September 2017

Design and Analysis of an Intelligent Speech Recognition System

[1]Baibaswata Mohapatra [2]Akanksha Sehgal
[1,2]Department of Electronics and Communication Engineering, Galgotias University, Yamuna Expressway Greater Noida, Uttar Pradesh
[1]bmohapatra@Galgotiasuniversity.edu.in, [2]akanksha.sehgal@Galgotiasuniversity.edu.in

Abstract: It is critical to recognize speech acknowledgment from speech comprehension (or speech distinguishing proof), the importance of an expression instead of its translation. Speech acknowledgment is likewise not the same as voice acknowledgment: though speech acknowledgment alludes to the capacity of a machine to perceive the words that are verbally expressed (i.e., what is said), voice acknowledgment includes the capacity of a machine to perceive talking style (i.e., who said something). The proposed research work is a savvy speech acknowledgment framework and it depends on Deep learning. The utilization of voice as a characteristic and supportive method for human-device correspondence is prevalently identified with sans hands things and correspondence with little structure factor gadgets. This new territory of AI has yielded far superior outcomes when contrasted with others in an assortment of utilizations including speech, and along these lines turned into an alluring zone of research. It is worked for train the framework the continuous condition and perform better at both raucous and raucous free condition having individuals of various talking styles and talking rate.

Keywords*:* Deep learning algorithm, Talking rate, Neural network, Talking style, Speech recognition.

INTRODUCTION

Since the most recent decade, Deep learning has emerged as another appealing region of AI, and as far back as has been inspected and used in a scope of various research themes. Deep learning comprises of a numerous of AI calculations nourished with contributions to the type of various layered models. These models are generally neural systems comprising of various degrees of non-direct activities. The AI calculations endeavor to gain from these deep neural systems by removing explicit highlights and data [1]. Preceding 2006, looking through Deep design inputs was not an anticipated straight forward undertaking; however, the improvement of Deep learning calculations helped to resolve this issue and disentangled the way toward looking through the parameter space of deep models.

Deep learning models can likewise work as an avaricious layer-wise unaided pre-processing. This implies it will take in the chain of command from extricated highlights from each layer at once. Highlight learning is accomplished via preparing each layer with an unaided learning calculation, which takes the highlights separated from the past layer and uses it as a contribution for the following layer. In this way, highlight learning will endeavor to

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 9, September 2017

get familiar with the change of the recently learned highlights at each new layer. Every cycle highlight learning adds one layer of loads to a deep neural system. So the layers with learned loads can be stacked to instate a Deep managed indicator [2].

Utilizing Deep structures has demonstrated to be progressively proficient in speaking to non-straight capacities in contrast with shallower models. Studies have demonstrated that fewer parameters are required to speak to a certain non-straight capacity in Deep engineering in examination with the huge number of parameters expected to speak to a similar capacity in a shallower design. This shows further structures are progressively productive from a measurable perspective. SSR (Smart Speech Recognition) frameworks encourage a physically crippled individual to the direction and control a machine. Indeed, even customary people would incline toward a voice interface over a console or mouse. The favorable position is increasingly evident if there should arise an occurrence of little handheld gadgets. The transcription machine is a notable use of SSR [3]. Because of the pervasive media transmission frameworks, speech interface is exceptionally helpful for information passage, access of data from remote databases, intuitive administrations, for example, ticket reservation. SSR frameworks are practical in situations where hands and eyes are occupied, for example, driving or medical procedures. They help show phonetic and modified instructing too. Figure 1 shows the architecture of SSR.

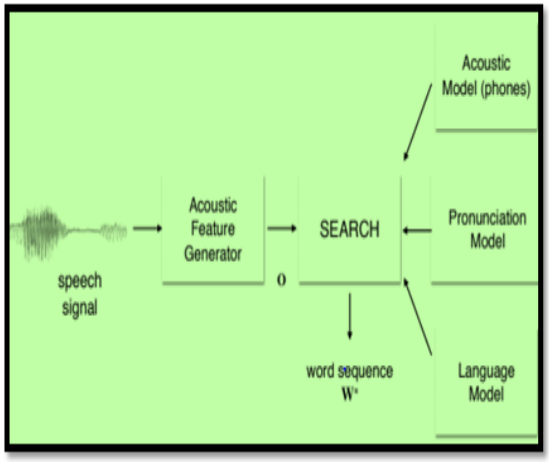

Figure 1: SSR Architecture

Human preference for speech communication has directed to the progress of different vocalized languages for discussion and information exchange. Such classifications need the development of a Reliable and multipurpose speech recognition system at its front-end, proficient in interpreting the spoken words. The primary objective of the work includes the development of Spontaneous Speech recognizer for multilingual systems which involves the generation of an acoustic model for multilingual speech, Phonetic Decoding, Language Modelling and also for the development of the speech corpus for training and testing purpose. The research work includes the step by step process for training the multilingual speech model. And also involves the test it on the live environment using different speakers who have a different accent, speaking styles [4]. Because of the spontaneous speech system, the sounds are usually unprompted and not- designed and are commonly described by repetitions, repairs, false start, partial words and non-planned words, silence gap, etc. In this research work, the focus is on the development of the spontaneous speech model for the recognition of the multilingual systems. So far, no

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 9, September 2017

work has been done for spontaneous speech recognition for the multilingual system (Punjabi, Hindi, and English) using the deep learning neural network [5]. To build up the Spontaneous Speech model for the multilingual system, a very dataset from the English, Hindi and Punjabi language has to be taken from the presentations, live debates, interviews, and telephonic conversations and one to one communications of the human being.

For the training and testing, Deep learning will be used to make the system more robust and reliable for recognition in the live environment. The real-time training will be done using a deep learning approach to improve the recognition accuracy of the proposed system. Another objective of the proposed solution is that it will be speaker impendent and trained with both male and female voices of all the mentioned languages for multilingual compatibility [6]. Because speaker-independent voice recognition is the opposite of speaker-dependent voice recognition. It does not need any training by the speaker and can recognize the speech from a large variety of speakers. More processing capability is required by Speaker independent voice recognition systems as compared with speaker-dependent systems. The interest on Multilingual Systems stimulates because there are common languages of Punjab (English, Punjabi, and Hindi) and two more frequently used languages in India (Hindi and English) are used. The performance of the proposed model will also be evaluated by using the speech recognition accuracy in both noisy and noise-free environment. Other parameters will also be computed such as word error rate, convergence ratio and overall likelihood per frame [7].

LITERATURE SURVEY

These early SSR frameworks utilized layout based acknowledgment concerning design coordinating that contrasted the speaker's information and pre-stored acoustic formats or examples. Example coordinating works well at the word level for acknowledgment of phonetically particular things in little vocabularies, however, it is less successful for bigger jargon acknowledgment [8]. Another confinement of example coordinating is its failure to coordinate and adjust input speech signals with presorted acoustic models of various lengths. Consequently, the exhibition of these SSR frameworks was dreary in light of the fact that they utilized acoustic methodologies that lone perceived fundamental units of speech obviously articulated by a solitary speaker [9].

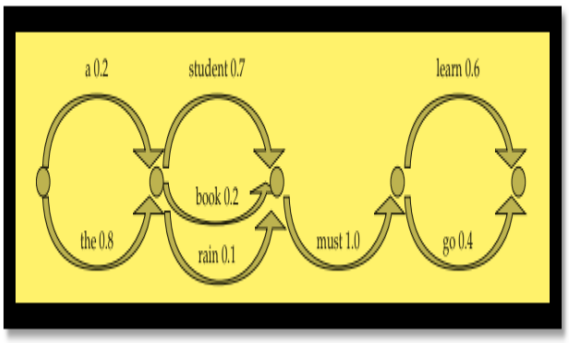

Figure 2: The Four State Markov Model

An early endeavor to build speaker-autonomous recognizers was first to utilize a PC. Afterward, analysts explored different avenues regarding time normalization methods, (for example, dynamic time traveling, or DTW) to limit contrasts in speech paces of various talkers and to dependably identify speech and closures. HMM depends on complex measurable and probabilistic investigations. In basic terms, shrouded Markov models speak to language units (e.g., phonemes or words) as an arrangement of states with change probabilities between each state. To move to start with one state then onto the next, the model will utilize the maximum conversion

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 9, September 2017

probability as shown in figure 2. The primary quality of HMM is that it can depict the likelihood of states and speak to their request and fluctuation through coordinating systems. HMM can sufficiently examine both the transient and ghostly varieties of speech flags, and can perceive and efficiently disentangle nonstop speech input [10].

METHODOLOGY

Information on the production of different speech sounds will assist us with understanding unearthly and worldly properties of speech sounds. This, thus, will empower us to describe sounds as far as highlights which will help in acknowledgment and order of speech sounds. Sounds are produced when air from the lungs energizes the air hole of the mouth. Figure 3 shows the message encoding and decoding phenomenon between the body and the system.

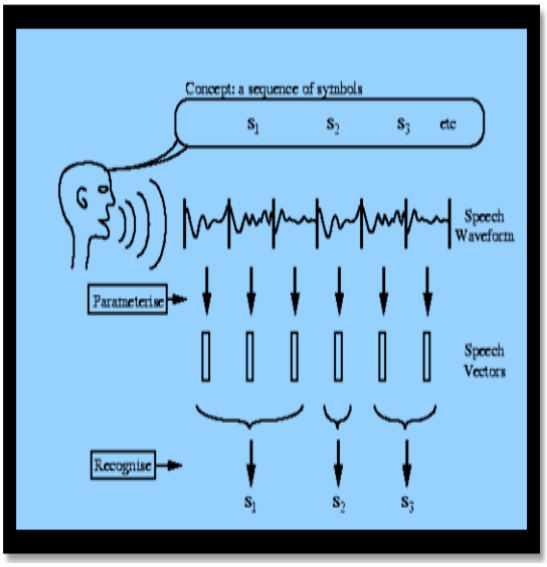

Figure 3: Message Encoding and Decoding

The proposed system first acquire the corpus from different sources after that train the corpus and create the dictionary file, extract the phones from the dictionary file as well as a unique word from the transcription. And perform the recording for the same transcription than apply the dynamic feature extraction mechanism to collect data from these files. After collecting data apply deep learning neural network for the generation and classification of robust spontaneous speech model for multilingual speech system [11]. The block diagram representation of the proposed method is shown in figure 4.

RESULTS AND CONCLUSION

However, advancement has been made in recent decades in the zone of communication in language innovation. This has prompted the sending of speech acknowledgment frameworks in a couple of utilization areas. However, the present designing models of speech and language don't satisfactorily show the regular language capacities of the human cerebrum. The psychological parts of the human cerebrum are intricate and the advancement of proper models is as yet a difficult research task. Such advancement will prompt Ubiquitous Speech Communication Interfaces through which individuals will have the option to collaborate with machines as helpfully and normally as do among themselves. The proposed system uses deep learning neural networks for the generation and classification purposes and checks the performance of the proposed model in a live environment. And computed the performance using parameters such as recognition accuracy, word error rate, convergence ratio and overall likelihood per frame. This system is also very useful and economical.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 9, September 2017

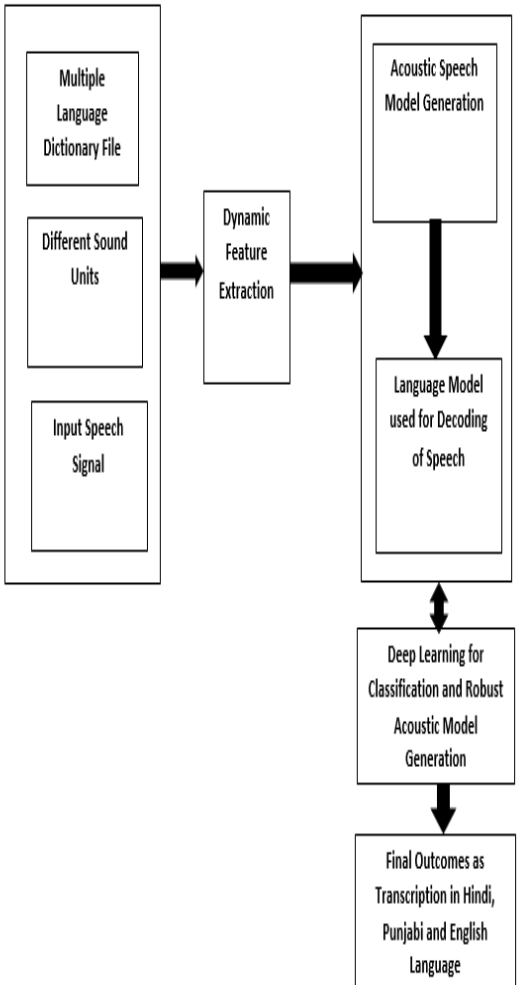

Figure 4: Block Diagram

REFERENCES

[1] S. Toshniwal et al., "Multilingual Speech Recognition with a Single End-to-End Model," in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2018, doi: 10.1109/ICASSP.2018.8461972.

[2] T. Schultz and K. Kirchhoff, Multilingual Speech Processing. 2006.

[3] J. B. Mariño, A. Moreno, and A. Nogueiras, "A first experience on multilingual acoustic modeling of the languages spoken in Morocco," in 8th International Conference on Spoken Language Processing, ICSLP 2004, 2004.

[4] R. Dufour, Y. Estève, and P. Deléglise, "Characterizing and detecting spontaneous speech: Application to speaker role recognition," Speech Commun., 2014, doi: 10.1016/j.specom.2013.07.007.

[5] N. T. Vu, Y. Wang, M. Klose, Z. Mihaylova, and T. Schultz, "Improving SSR performance on non-native speech using multilingual and crosslingual information," in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2014.

[6] S. Watanabe, T. Hori, and J. R. Hershey, "Language independent end-to-end architecture for joint language identification and speech recognition," in 2017 IEEE Automatic Speech Recognition and Understanding Workshop, SSRU 2017 - Proceedings, 2018, doi: 10.1109/SSRU.2017.8268945.

[7] H. Tang, W. Liu, W. L. Zheng, and B. L. Lu, "Multimodal Emotion Recognition Using Deep Neural Networks," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017, doi: 10.1007/978-3-319-70093-9_86.

[8] W. Song and J. Cai, "End-to-End Deep Neural Network for Automatic Speech Recognition," CS224N Proj., 2015.

[9] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, "Speech Recognition Using Deep Neural Networks: A Systematic Review," IEEE Access, vol. 7, pp. 19143–19165, 2019, doi: 10.1109/ACCESS.2019.2896880.

[10] Samudravijaya K, "Automatic Speech Recognition."

[11] J. PECKHAM, "AUTOMATIC SPEECH

RECOGNITION.," NEW ELECTRON, vol. V 15, no. N 18, 1982, doi: 10.1002/9781405198431.wbeal0066.

.

[12] Vishal Jain, Dr. Mayank Singh, "Ontology Based Web Crawler to Search Documents in the Semantic Web", "Wilkes100 - Second International Conference on Computing Sciences", in association with International Neural Network Society and Advanced Computing Research Society, held on 15th and 16th November, 2013 organized by Lovely Professional University, Phagwara, Punjab, India and proceeding published by Elsevier Science.

[13] Vishal Jain, Dr. Mayank Singh, "Ontology Development and Query Retrieval using Protégé Tool", International Journal of Intelligent Systems and Applications (IJISA), Hongkong, Vol. 5, No. 9, August 2013, page no. 67-75, having ISSN No. 2074-9058, DOI: 10.5815/ijisa.2013.09.08 .

[14] S.Balamurugan , L.Jeevitha, A.Anupriya and Dr.R.Gokul Kruba Shanker, "Fog Computing: Synergizing Cloud, Big Data and IoT- Strengths, Weaknesses, Opportunities and Threats (SWOT) Analysis", International Research Journal of Engineering and Technology (IRJET), Volume 3 issue 10, e-ISSN: 2395 -0056, p-ISSN: 2395-0072, 2016

[15] S.Balamurugan, S.Dharanikumar, D.Gokul Prasanth, Krithika, Madhumitha, V.M.Prabhakaran and Dr.R.Gokul Kruba Shanker, "Internet of Safety: Applying IoT in Developing Anti Rape Mechanism for Women Empowerment", International Research Journal of Engineering and Technology (IRJET), Volume 3 issue 10, pp.713-719, e-ISSN: 2395 -0056, p-ISSN: 2395-0072, 2016