Machine Learning Techniques for Speech Recognition using the Magnititudes

Angeline Valentina Sweety A., Gopala Krishnan C., Mukesh Krishnan M.

Abstract: ***Speech is the most proficient method of correspondence between people groups. Discourse acknowledgment is an interdisciplinary subfield of computational phonetics that creates approaches and advances that empowers the acknowledgment and interpretation of communicated in language into content by PCs. It is otherwise called programmed discourse acknowledgment (ASR), PC discourse acknowledgment or discourse to content (STT). It consolidates information and research in the etymology, software engineering, and electrical building fields. This, being the best methodology of correspondence, could likewise be a helpful interface to speak with machines. Machine learning consists of supervised and unsupervised learning among which supervised learning is used for the speech recognition objectives. Supervised learning is that the data processing task of inferring a perform from labeled coaching information. Speech recognition is the current trend that has gained focus over the decades. Most automation technologies use speech and speech recognition for various perspectives. This paper offers a diagram of major innovative point of view and valuation for the fundamental advancement of speech recognitionand offers review method created in each phase of discourse acknowledgment utilizing supervised learning. The project will use ANN to recognize speeches using magnitudes with large datasets.***

Keywords—Deep Neural Network, Speech Recognition, Magnitude

I. INTRODUCTION

Speech Recognition which also called Automatic Speech Recognition (ASR) perceives the expressed words and people and changes over them to a machine-meaningful configuration. By changing over spoken sound into content, speech recognition innovation improves the clients to control advanced gadgets. Rather than utilizing regular devices, for example, keystrokes, catches, consoles and so forth. The ASR technique can be broadly used in all walks of life [1]. In basic words speech recognition can be characterized as the way toward changing over speech signal to a succession of words by methods for calculation actualized as a PC program. Speech process is one among the energizing regions of sign procedure. The point of speech recognition is to build up a method for speech contribution to machine based readable script, which can be very well used in libraries, banks, and various workplaces for effective management.

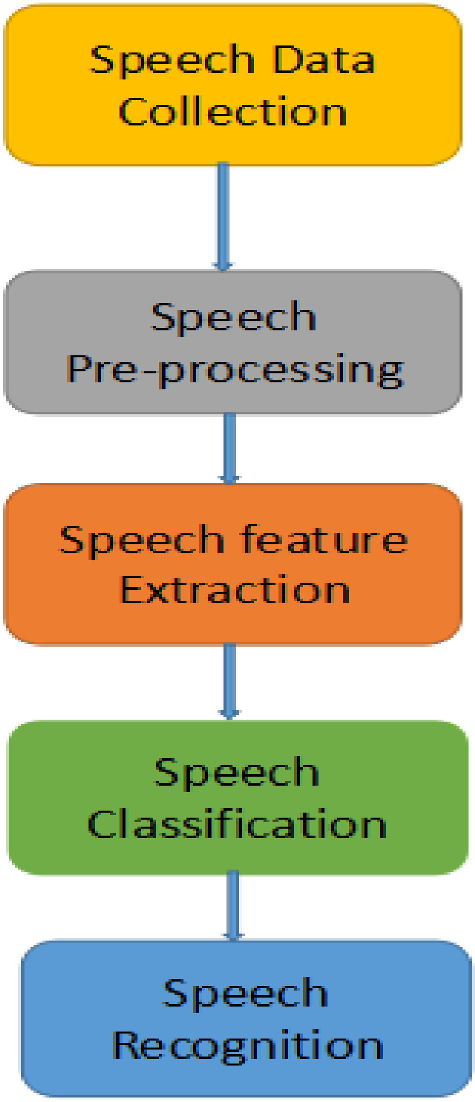

Fig : Speech Recognition

Automatic speech recognition today finds across the board application in assignments that require human machine interface, for example, programmed call handling.

Revised Manuscript Received on February 15, 2020.

* Correspondence Author

Angeline Valentina Sweety A., Department of Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli, India. Email:valentinaraj1997@gmail.com

Dr. Gopala Krishnan C., Department of Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli, India. Email: skywarekrish@gmail.com

Mukesh Krishnan M., Department of Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli, India.

II. EXISTING SYSTEM

In the current framework, a librarian can use barcode to get a particular book details and its availability in the library. But the exact location where the book is placed cannot be easily identified. It is very difficult and laborious to get a particular book or books when large volume of books, manuals, research papers etc. are stacked in big libraries. Moreover it is time consuming and there is every possibility for misplacement and wrong issue of books or journals etc.

III. PROPOSED SYSTEM

In the existing system various algorithms are used whereas in the proposed system we have implemented only Artificial Neuron Network (ANN) and accordingly input data's have to be fed and output results probably the exact one is expected to be derived. By this means considerable time saving is expected and financial implications too can be minimized.

A. Automatic Speech Recognition (ASR):

Fundamental Principle:

ASR systems operate in two phases. First, a training part, throughout that the system learns the reference patterns representing the various speech sounds (e.g. phrases, words, phones) that represent the vocabulary of the applying.

Each reference is learned from spoken examples and hold on either within the sort of templates obtained by some averaging technique or models that characterize the statistical properties of pattern. Second, a recognizing phase, during which an unknown input pattern, is identified by considering the references.

B. Speech Recognition Techniques:

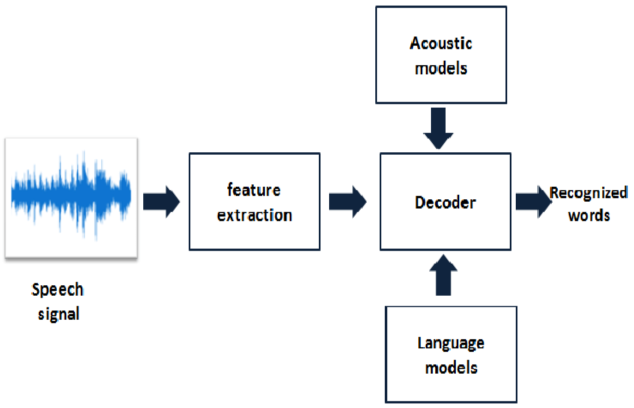

Fig: Speech Recognition Techniques

The objective of Speech Recognition is for a machine to be prepared to "hear," see," and "follow up on" verbally expressed information. The soonest discourse acknowledgment frameworks were first attempted inside the mid-1950s at Bell Laboratories. The objective of Automatic Speech Recognition is to examine, extricate portray and recognize information concerning the speaker character. The speaker framework is likewise seen as working during a four phases

a. Assessment
b. Feature extraction
c. Demonstrating
d. Testing

IV. IMPLEMENTATION

A. Neural network

Neural systems have numerous similitudes with Markov models. Both are factual models that are spoken to as charts. Where Markov models use conceivable outcomes for state advances, neural systems use affiliation qualities and capacities. A key differentiation is that neural systems are essentially parallel though Markov chains are sequential. Frequencies in speech happen in parallel, while syllable neural systems have numerous similitudes with Markov models. Both zone unit measurable models that are envisioned as charts. Where Markov models use chances for state changes, neural systems use association qualities and capacities.

B. Artificial Neuron

Artificial Neurons are the fundamental unit of Artificial Neural Network that simulates the four basic perform of biological nerve cell. It is a scientific capacity imagined as a model of common neuron. The following figure shows the basic artificial neuron.

Types of Artificial Neural Network

- Feedforward Network
- Recurrent Neural Network
- Modular Neural Network
- Kohonen Self Organizing Maps

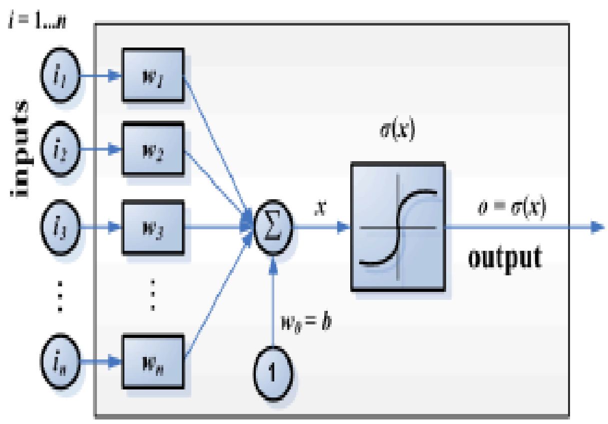

Fig :Artificial Neuron

C. Feedforward Network

A Feed forward Network is a computational model dependent on the structure and elements of natural neural systems. Data that courses through the system influences the structure of the ANN on the grounds that a neural system changes - or learns, as it were - in view of that info and yield.

ANN's are viewed as nonlinear factual information demonstrating devices where the perplexing connections among sources of info and yields are displayed or designs are found. ANN is otherwise called a neural system. ANN's have three layers that are interconnected. The main layer comprises of input neurons. Those neurons send information on to the subsequent layer, which thus sends the yield neurons to the third layer [4].

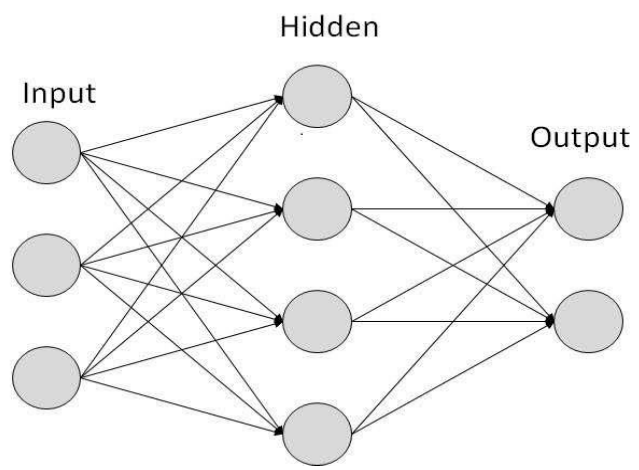

Fig :Artificial Neural Network

Artificial neural networks are an endeavor to get PCs to work progressively like the human brain. Your brain doesn't store explicit encoded directions; it's enormous systems of neurons. These change their associations with each option as new data past through them. Speech recognition abuse this AI is jumping out at firms like Google and Microsoft, who have tremendous databases of information to mentor these systems.

An ANN has a few preferences however one of the most perceived of these is the way that it can really gain from watching informational indexes. As such, ANN is utilized as an irregular capacity estimation device. These types of devices encourage gauge the premier productive and perfect procedures for inbound at arrangements though molding processing capacities or appropriations. ANN takes information tests rather than whole information sets to pick up arrangements, which sets aside both time and cash. ANN's zone unit contemplated genuinely clear scientific models to support existing information investigation innovations.

D. Working with Audio Files

Speech Recognition makes working with sound records simple gratitude to its helpful AudioFile class. These categories are often initialized with the trail to AN audio file and provide a context manager interface for reading and dealing with the file's contents.

Strengthening File Types Currently, Speech Recognition supports the following file formats:

●WAV: must be in PCM/LPCM format

●AIFF

●AIFF-C

●FLAC: must be native FLAC format; OGG-FLAC is not stronger

Recognizing speech requires sound information, and Speech Recognition makes recovering this info extremely simple. Rather than making contents for getting to mouthpieces and procedure sound documents without any preparation, Speech Recognition can have you ever ready for action in just number of minutes.

The adaptability and usability of the Speech Recognition bundle settle on it a magnificent decision for any Python project. Be that as it may, support for each element of each apus it wraps isn't verified. You can must be constrained to pay some time investigating the offered decisions to search out if speech Recognition will include your particular case.

E. Optimization

At the point when speech acknowledgment is being created, the most intricate issue is to make search exact and to make it sufficiently quick to not run for a very long time. Since models aren't great, another test is to make the model match the speech.

Word Error Rate: How about we expect we have a unique book and acknowledgment content with a length of N words. I is the quantity of embedded words, D is the quantity of erased words and S speak to the quantity of subbed words. With this, the word blunder rate can be determined as

$$WER = (I + D + S)/N$$

The WER is generally estimated in percent.

Table: WER and Accuracy

Model	WER	Accuracy
Bigram	50.69	73.4
Context-dependent	21.59	12.5
Unigram	11.34	87.5

Accuracy: It is nearly equivalent to the word mistake rate, yet it doesn't consider.

$$Accuracy = (N - D - S)/N$$

For most errands, the precision is a more regrettable measure than the WER, since inclusions are additionally significant in the conclusive outcomes. Be that as it may, for certain undertakings, the precision is a sensible proportion of the decoder execution.

Speed: Assume a sound document makes some account memories (RT) of 2 hours and the unraveling took 6 hours. At that point the speed is considered 3xRT.

V. TRAINING AN ARTIFICIAL NEURAL NETWORK

Artificial neural systems are nearly unrefined electronic systems of "neurons" strengthened the structure of the brain. They process records each in turn, and "learn" by looking at their characterization of the record (which, at the start, is to a great extent subjective) with the known real grouping of the record. The blunders from the underlying characterization of the essential record are nourished back to the system, and wont to change the systems algorithmic program the second time around, etc for some constant.

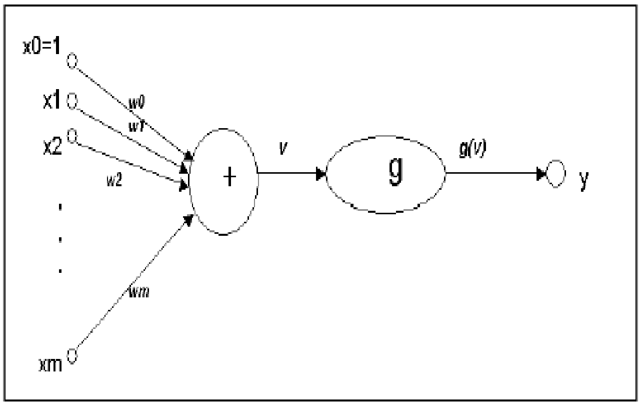

Fig: Training an artificial neural network

Generally, a neuron in an ANN is
1. A assortment of input gain (xi) and related loads (wi).
2. A work (g) that amount the loads and maps the outcomes to a yield (y).

In the training part, the right category for every record is thought (this is termed supervised training), and also the output nodes will therefore be assigned "correct" values-"1" for the node similar to the right category and others as "0".

VI. RESULT

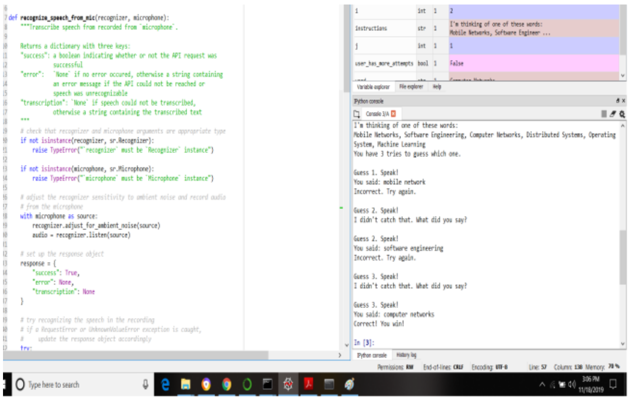

Fig : Working with Audio Files

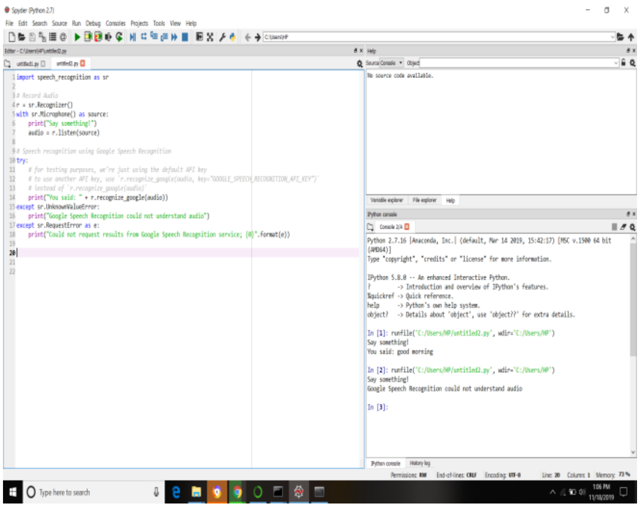

Fig : Extraction of recognized speech

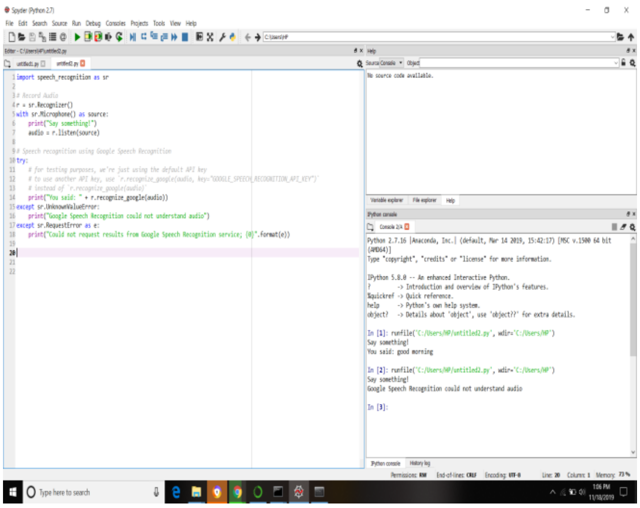

Fig: Recognized Speech Outcomes

VII. CONCLUSION

In this paper, we have tried to introduce a simple technique which could be used to recognize connected speech and the person concerned. The speech features extracted are compared with already saved speeches in the database for matching. This procedure makes it feasible by using the speech of the presenter and it will be easy to authenticate their individuality. It gives control access to different applications, for example, google discourse, internet business, window speech identification, m-trade, vehicle mechanization, home robotization and security control and so forth. Indeed the application of this technique will certainly enhance smooth and perfect administrative innovations in the day today activates wherever manpower is entertained in multiples such as libraries banks and various workplaces etc. In the present study I have tried to develop a device which will enable to find the presence of a particular data from the cluster of dataset using python. In future this device can be taken to the next level by using Artificial Neural Network (ANN) which will lead us to work with incomplete knowledge on information's related to the speech and the person concerned. Further in the new application the network layers will be built and trained to show the pictorial representation of in-built data.

REFERENCES

1. Ali bou nassif, et al.(2019), "Speech Recognition Using Deep Neural Networks: A Systematic Review" Digital Object Identifier 10.1109/ACCESS.2019.2896880.
2. Bhushan C. Kamble (2016), "Speech Recognition Using Artificial Neural Network –A Review"Int'l Journal of Computing, Communications & Instrumentation Engg. (IJCCIE) Vol. 3.
3. Chen M. and Zhou P. and G. Fortino(2016),"Emotion communication system", *IEEE Access*, vol. 5, pp. 326337.
4. C. G. Krishnan, K. Sivakumar and E. Manohar, "An Enhanced Method to Secure and Energy Effective Data Transfer in WSN using Hierarchical and Dynamic Elliptic Curve Cryptosystem," 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2018, pp. 1-7.
5. C. Krishnan, A.Rengarajan and R.Manikandan, "Delay Reduction by Providing Location Based Services using Hybrid Cache in Peer to Peer Networks," KSII Transactions on Internet and Information Systems, vol. 9, no. 6, pp. 2078-2094, 2015. DOI: 10.3837/tiis.2015.06.006.
6. Deng J. and Xu X. and Zhang Z. and Frühholz S. and Schuller B.(2017),"Universum auto encoder-based domain adaptation for speech emotion recognition",' *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 500504.
7. Fayek H. M and Lech .M and Cavedon .L (2017),"Evaluating deep learning architectures for speech emotion recognition"," Neural Netw., vol. 92, pp. 6068.
8. Hossain M. S. and Muhammad G.(2019), "Emotion recognition using deep learning approach from audiovisual emotional big data"," *Inf. Fusion*, vol. 49, pp. 6978.
9. Hynek Hermansky (2013),"Multistream Recognition of Speech: Dealing With Unknown Unknowns" Proceedings of the IEEE | Vol. 101, No. 5. *IEEE Int. Conf.* Oct. 2017.
10. Javier Gonzalez-Dominguez and David Eustis and Ignacio Lopez-Moreno and Andrew and Françoise Beaufays and Pedro J. Moreno (2015), "A Real-Time End-To-End Multilingual Speech Recognition Architecture", IEEE journal of selected topics in signal processing, vol. 9, no.4.
11. Le D. and Provost. E. M (2013), "Emotion recognition from spontaneous speech using hidden MARKOV models with deep belief networks"," in *Proc. IEEE Workshop Autom. Speech Recognit. Understand.*,pp. 216221.

12. Naijun Zheng and Xiao-Lei Zhang (2019), "Phase-Aware Speech Enhancement Based on Deep Neural Networks" IEEE/ACM Transactions On Audio, Speech, And Language Processing, VOL. 27, NO.1
13. P.S.Apirajitha, C.Gopala Krishnan, G.Aravind Swaminathan, E.Manohar(2019) "Enhanced Secure User Data on Cloud using Cloud Data Centre Computing and Decoy Technique", "International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-9, July 2019"
14. Rázuri J. G. And Sundgren D. and Rahmani R. and Moran A. and Bonet I. and Larsson A.(2015), "Speech emotion recognition in emotional feedback for human-robot interaction"," *Int. J. Adv. Res. Artif. Intell.*, vol. 4, no. 2, pp. 2027.
15. Santosh Gaikwad K. and Bharti Gawali W. and Pravin Yennawar (2010), "A Review on Speech Recognition Techniques", IJCA Vol. 10, No. 3, pp. 16-24.
16. Yashwanth H and Harish Mahendrakar and Suman Davia (2004),"Automatic Speech recognition Using Audio Visual Cues", IEEE India Annual Conference pp. 166-169.
17. Zhang S. and Huang T., and Gao W.(2017), "Speech emotion recognition using deep convolutional neural network and discriminant temporal pyramid matching","*IEEE Trans. Multimedia*, vol. 20, no. 6, pp. 15761590.

AUTHORS PROFILE

Angeline Valentina Sweety A., received the B.E degree in Computer Science and Engineering from Franics Xavier Engineering College, Affiliated to Anna University, Tirunelveli, Tamil Nadu, India in 2014. She is currently pursuing the M.E degree in Computer Science and Engineering with Francis Xavier Engineering College, An Autonomous Institution, Tirunelveli ,Tamil Nadu, India.

Dr. Gopala Krishnan C., Professor, Department of Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli, Tamil Nadu, India. He received his B.E., M.E., and Ph.D degrees in Computer Science and Engineering. His Academic Research was Cloud Computing. His Research Interests are Mobile Communication and Cloud Computing. He has fifteen publications. He has attended three seminars, twenty conferences, nine workshop and has delivered a lecture. He is an active member of ISTE. He has applied for DST-SHRI Sponsored Research.

Mukesh Krishnan M., Professor, Department of Computer Science and Engineering, Francis Xavier Engineering College, Tirunelveli, Tamil Nadu, India. He received his B.E., M.E., degrees in Computer Science and Engineering. His Academic Research was Cloud Computing and Image Processing. His Research Interests are Image Processing and Cloud Computing. He has six publications. He has attended two seminars, two conferences, three workshop and has delivered a lecture. He is an active member of ISTE. He has applied for DST-SHRI Sponsored Research.