iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.42237

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Voice Isolation Using Artificial Neural Network

Vikash Kumar[1], Shivam Kumar Gupta[2], Harsh Sharma[3], Uchit Bhadauriya[4], Chandra Prakash Varma[5]
[1, 2, 3, 4]*Electronics Engineering, Dr. Ambedkar Institute Of Technology for Handicapped, Kanpur, UP*
[5]*Assistant Professor, Electronics Engineering, Dr. Ambedkar Institute of Technology for Handicapped, Kanpur*

Abstract: ***The paper reflects the use of Artificial Neural Networks with the help of various machine learning algorithms for voice isolation. In particular, we consider the case of a voice sample recognition by analyzing the speech signals with the help of machine learning algorithms such as artificial neural networks, independent component analysis, activation function. The strategies by which our central nervous network decodes the network stimuli same as artificial neural network will analyze the given speech sample. After first step, a set of machine learning algorithms will be used like independent component analysis algorithm and gradient function algorithm for processing. After the processing, a decision statement will be applied to generate the desired output.***

Keywords: Artificial Neural Network, Voice Isolation, Fast Independent Component Analysis, Gradient Descent

I. INTRODUCTION

Vitality of the voice isolation has been observed recently especially in case of the mobile communication as now, the current mobile users stand at 800 million after the arrival of 4G before it the mobile user base was roughly around only 25 million and these numbers will keep on increasing. The arrival of 4G was definitely a boon for the people of India but as every coin has two sides it has its consequences like poor voice quality, e-waste management etc. This paper is solving one of many problems and i.e., voice isolation using artificial neural networks. Here, some mixed speech samples will be taken and different voices in the samples will be isolated from each other with the help of artificial neural network and some machine learning algorithms like Independent Component Analysis algorithm and Gradient Descent algorithm to get the desired output.

ANN works just like biological nervous system where one's output is input for another node. It consists of mainly 3 layers i.e., input layer, hidden layer and output layer. Artificial Neural Network (ANN) will be used for analyzing the sample speech signal. ICA and GD (Gradient Descent) algorithm used to separate the independent sources from the mixed speech signal. The activation function as a function which defines the output of a node when we give set of input signals to it. Then this input will work as an output for next node.

II. LITERATURE SURVEY

This phenomena was referred to be the ability to select and isolate one source of audio in a noisy environment from the others so that it can be listened to efficiently. To solve this problem numerous efforts have been made in many fields be it neurobiology, physiology, computer science or engineering. The deep neural network is based on the concept of deep learning that is the sub section of the machine learning and utilizes the facets of the artificial intelligence to classify and order the data's, the following section provides the deep neural network architectures employed in the various areas providing a state of art accuracy.

In 2011, Dr. R.L.K. Venkateswarlu, Dr. R. Vasantha Kumari, G. VaniJayaSri utilizes Recurrent Neural Network, one of the Neural Network techniques to observe the difference of alphabet from E- set to AH – set. In their research 6 speakers (a mixture of male and female) are trained in quiet environment. The English language offers and reumber of challenges for speech recognition. They used multilayer back propagation algorithm for training the neural network. Six speakers were trained using the multilayer perceptron with 108 input nodes, 2 hidden layers and 4 output nodes each for one word, with the noncurvear activation function sigmoid. The learning rate was taken as 0.1, momentum rate was taken as 0.5.Weights were initialized to random values between +0.1 and -0.1 and accepted error was chosen as 0.009. They have compared the performance of neural network with Multi-Layer Perceptron and concluded that RNN is better than Multi-Layer Perceptron. For A-set the maximum performances of speakers 1-6 were 93%, 99%, 96%, 93%, 92% & 94%. For E-set it was 99%, 100%, 98%, 97%, 97% & 95%, and For EH-set 100%, 95%, 98%, 95%, 98% & 98% and lastly for AH-set 95%, 98%, 96%, 96%, 95% & 95% respectively. Results shows that RNN is very powerful in classifying the speech signals.

Song, W., &Cai, J. (2015) has developed end to end speech recognition using hybrid CNN and RNN. They have used hybrid convolutional neural networks for phoneme recognition and HMM for word decoding. Their best model achieved an accuracy of 26.3% frame error on the standard core test dataset for TIMIT.

Their main motto is to replace GMM-HMM based automatic speech recognition with the deep neural networks. The CNN they used consists of 4 convolutional layers. The first two layers have max pooling and the next two densely connected layers with a softmax layer as output. The activation function used was ReLu. They implemented a rectangular convolutional kernel instead of square kernel.

In 2017, Microsoft researchers reached a historical human parity milestone of transcribing conversational telephony speech on the widely benchmarked Switchboard task. Multiple deep learning models were used to optimize speech recognition accuracy. The speech recognition word error rate was reported to be as low as 4 professional human transcribers working together on the same benchmark, which was funded by IBM Watson speech team on the same task.

The next fifteen years have been very fruitful for the SR systems. The size of the vocabulary became infinite. Recognition rates also improved for real time speech recognition problems. The ASR is still considered as a standard classification problem. It can identify sequences of words from speech wave forms. However, it had several issues that prevented it from achieving the desired satisfactory performance. These include multi model recognition, multilingual recognition and noisy environment. Weiner filtering, spectral subtraction or windowing can be used for noise removal and enhancement of the speech. The Gaussian mixture model (GMM) and HMM are widely used for acoustic modeling of the speech inputs. The key technologies developed in the last decade were the ANN, Deep Neural Networks (DNN), etc. These advancements have provided the way for today's speech recognition systems with infinite vocabulary size along with spontaneous speech recognition. A relatively new classification technique of support Vector Machines for emotion recognition during ASR has been explained by authors. SVM's provide a low cost solution to the classification of high dimensional vectors. The speech recognition systems have grown from template matching to HMMs, from filter banks to cepstral features, from smaller vocabularies to large vocabularies, from the speaker dependent technology to speaker independent technology. Due to the shorter computation time than the other systems, the ANNs can be used to produce an avatar system with real time speech talks. These neural networks deliver superior performance. Although, it takes longer time to train it, particularly when the ANNs have multiple hidden layers. Moreover, the process through which the ANNs were initialized greatly affect the performance of these networks. Therefore, deep neural networks were preferred for large quantity of unlabeled data.

Today speech technology plays an important role in many applications. Speech technology has moved from research to commercial application. Many human machine interfaces have been invented and applied today in telephone food ordering system, telephone directory assistance, air port information system, ticketing system, restaurant reservation system, spoken database querying for novice users, "handsbusy" applications in medicine or fieldwork, office dictation devices, or even automatic voice translation into foreign languages etc. Investigation has shown that more than 85% of people are satisfied with the capability of the information inquiring service system of speech recognition (Jiang, 2009).

Automatic Speech Recognition (ASR) and spoken language understanding are one of the most important part of applied machine intelligence. In they have focused on isolated voice command recognition for autonomous man-machine and intelligent robotic systems, they have created grammar model for small testing of command set with self-loops for each state to return blank symbols for noise and out of vocabulary words. They have compared recognition accuracy and average decision-making time of our approach with the state of the art continuous speech recognition engines based on language models and it has been experimentally proved that their approach was achieved 60% higher accuracy than conventional offline speech recognition methods based on language models.

III. ARTIFICIAL NEURAL NETWORKS

ANN is inspired by the biological neural networks. It is basically an interface that allows different algorithms of machine learning to work together. There are many places where it works. E.g. - Image recognition, noise recognition and voice isolation. A group of interconnected nodes is what ANN is. Each node's output is the input of another node. The connection by which artificial neurons or nodes are connected are called edges. All neurons and edges are assigned a weight that adjusts accordingly as the learning continues. The Artificial Neural Network is broadly divided into three layers input layer, hidden layer and output layer. Each node of input layer is connected to every node of hidden layer. Similarly, each node of hidden layer is connected to every node of output layer.

A. Layers of Artificial Neural Network

1) *Input Layer:* All the inputs are given to the model with the help of this layer. The condition for which we are training the neural network should be represented in the input layer. Each input should show some independent variable so that they can have an effect on the output of network.

2) *Hidden Layer:* The data on which activation function is applied are collected and forms this hidden layer which lies in the middle of input layer and output layer. It does the processing achieved by the previous layer. A neural network can be consisted of many hidden layer depends upon how complex the problem is. If it can be separated linearly then it means activation function can be implemented to input layer and thus no hidden layer is required, where as if the decisions to be made is complex then 3 to 5 hidden layers can be used.
3) *Output Layer:* This layer makes data available after it has been processed. It collects and transmit data in the designed way. The pattern that the output layer tells can help one trace its route back to the input layer. Number of data parameters in the output layer describes the work that ANN is performing.

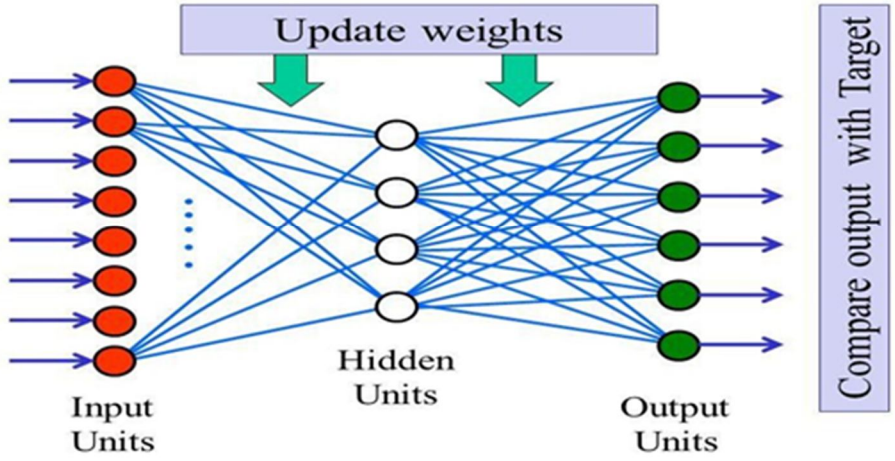


Fig. 1 Neural Network Architecture

IV. FAST INDEPENDENT COMPONENT ANALYSIS

Fast ICA is an effective and efficient algorithm for ICA. Fast ICA uses two important algorithms – gradient descent and fixed-point iteration. Fast ICA seeks an orthogonal rotation of pre whitened data, through a fixed-point iteration scheme, that maximizes a 32 measure of non-Gaussianity of the rotated components To apply this algorithm we need to pre-whiten the data[9]. Whitening is an important pre-processing strategy in ICA algorithm. That is before applying the algorithm we transform the observed vector or the mixing matrix into orthogonal matrix. We need to transform X such that its components are uncorrelated and thus variance is equal to unity. Or we can say co-variance matrix = identity matrix. In our project we will be implementing fast ICA on the source signal X by first creating weight vector and then compare to get proper whitened data.

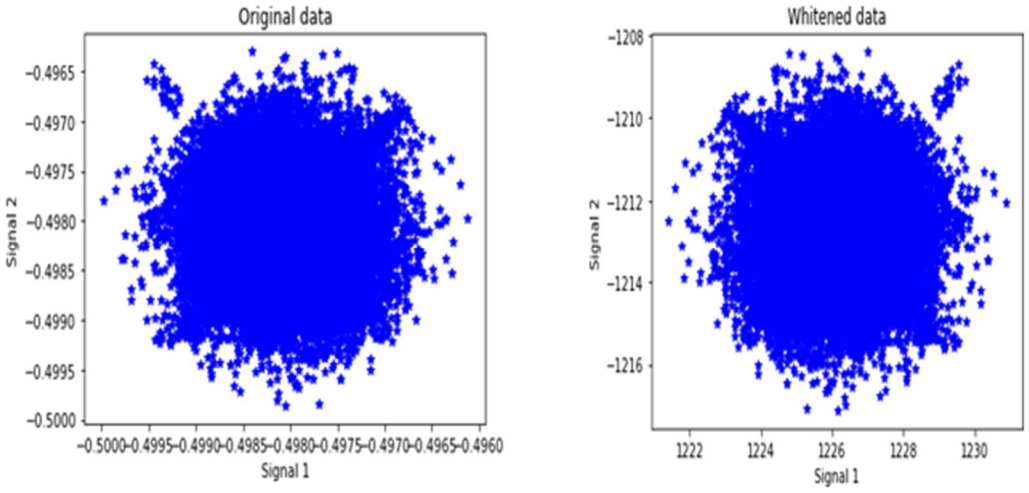


Fig. 2 Original And Whitened Data Graph

V. GRADIENT DESCENT

Gradient descent is an optimization algorithm which is used to minimize functions. Gradient descent is used to find the global minima of a function. Taking a real-life example suppose someone is standing in a valley he sees a point which appears to be the lowest point but that may not be the lowest point of the valley, thus we have two points one which appears to be the lowest and one which actually is. The one which appears to be is the local minima and the one which actually is, is the global minima. Gradient descent is an algorithm which help us to determine the global minima of a function.

The factor that affects the correctness of gradient descent is learning rate. With high learning rate we can cover more ground with the risk of overshooting the lowest point while with low learning rate we can move with confidence towards the global minima but it consumes more time.

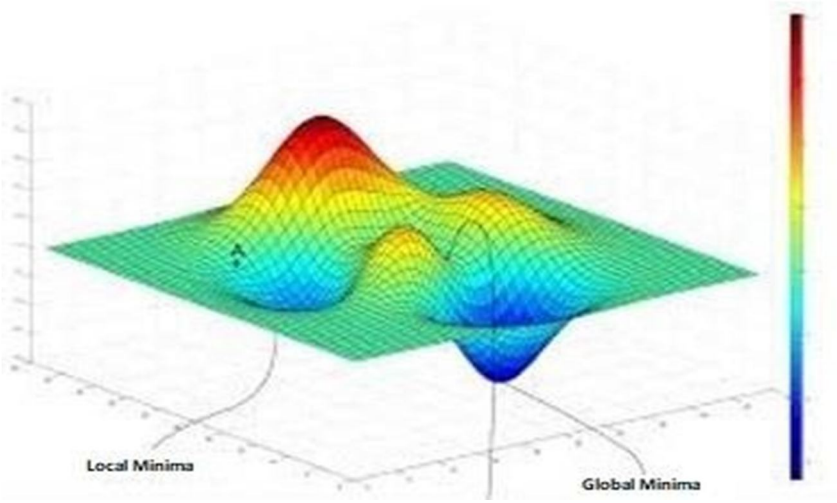


Fig. 3 Gradient Descent

VI. WORK SYSTEM

Voice Isolation is done by making efficient use of Independent Component Analysis Algorithm and Gradient Descent Algorithm. The methodology of this system can be divided into five steps. First, M sources are taken as input. In second step, these M sources are to be recorded at discrete interval of time by M distinct microphones. Then, the amplitude of the sound will be recorded and it will be compared with pitch of recorded sounds with the given frequencies and after that Activation function will be applied. In the fourth step, the data will be trained using epoch (The term epoch means training the neural network with all the training data for only one cycle.) to attain best result. After that ICA Algorithm and Mathematical conditions will be applied. Then, It checks a condition, if epoch is less than number of iterations. If the condition is true, new random parameters are passed along with input signals and ICA Algorithm is applied again. This process goes on until the condition is false. In the final step, this trained data will be used to separate the voice from mixed signals and generate different audio files.

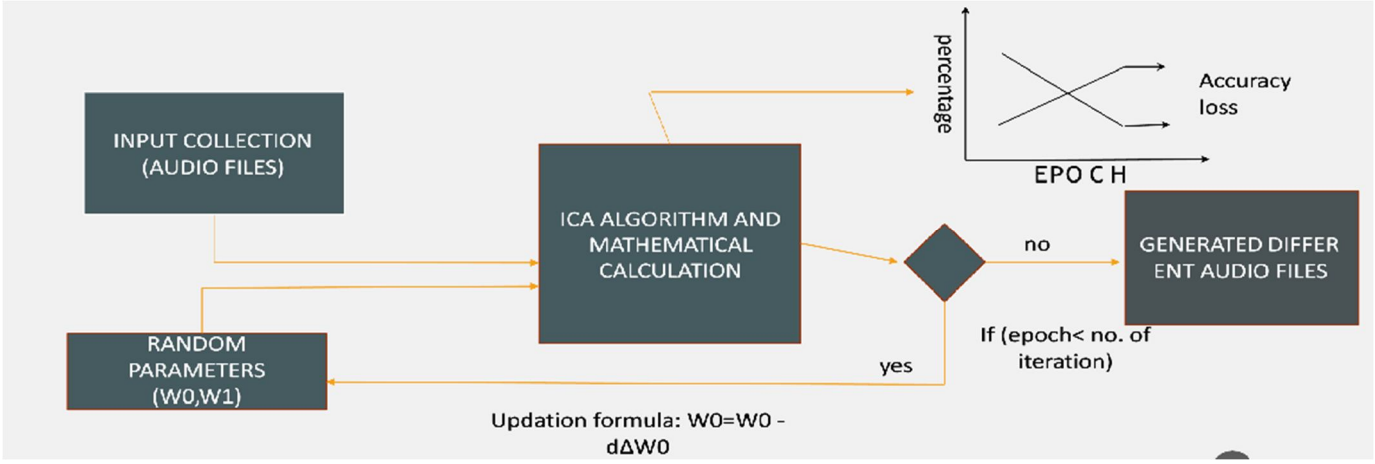


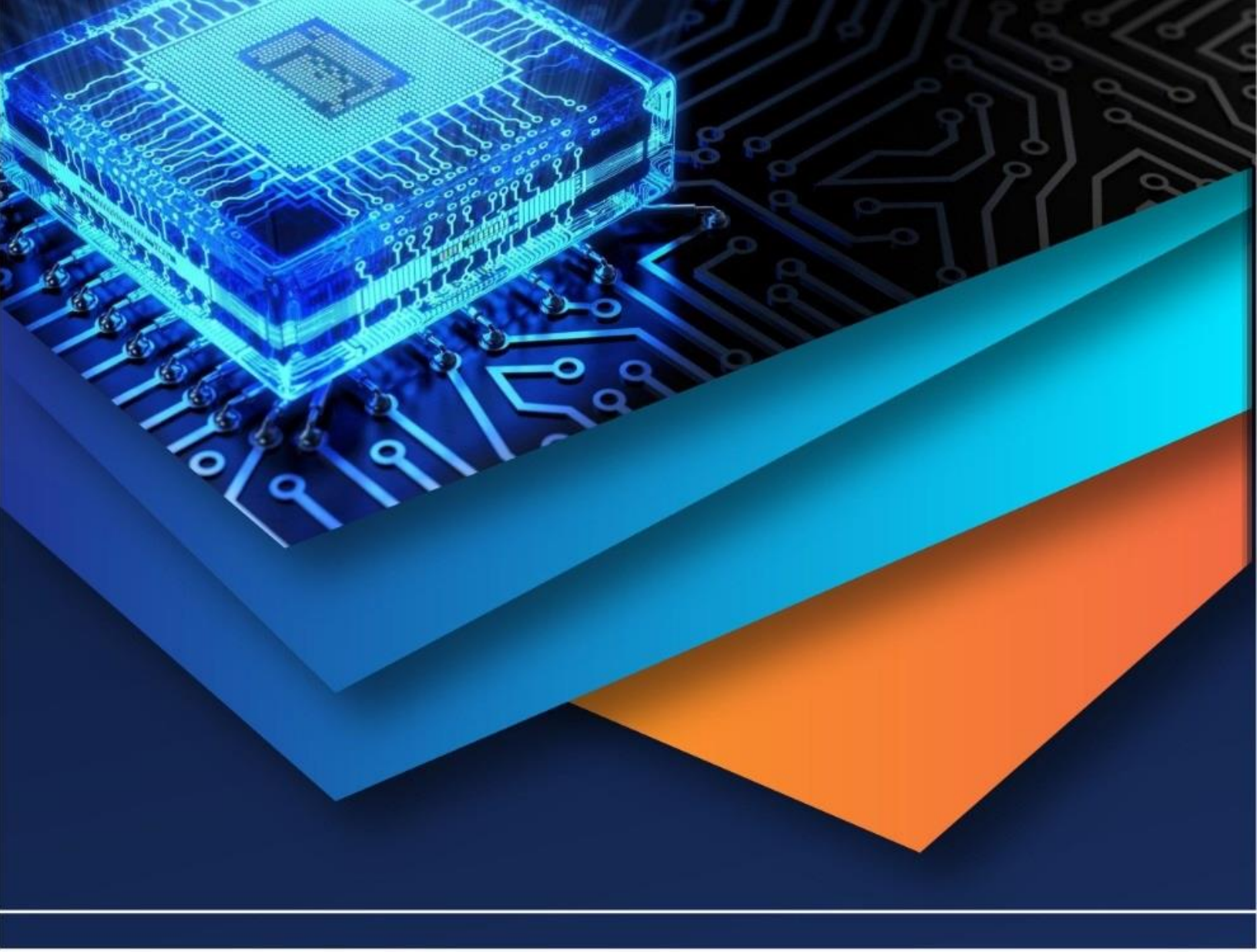
Fig. 4 Work Flow Diagram

VII. CONCLUSION & FUTURE SCOPE

The main objective of this project is to separate mixture of two voices or signals recorded via two microphones and two speakers into two individual project - the speech recording part and the computational part. Former part consists of audio mixtures sample, while latter, the computational part, consists of the algorithms and the computer programs that are used to separate the two voices and provide them as separate outputs. By implementing this project, we will be able to listen and understand one's voice properly without any noise in it and can easily understand what a person was saying. This project can also be used to isolate the sound of musical instruments and the vocals part of the singer and provide them as two different audio files.

REFERENCES

[1] https://ieeexplore.ieee.org/document/595489
[2] https://www.researchgate.net
[3] Independent Component Analysis, a new concept? Research done by Pierre Comon, 1994 (https://www.cs.purdue.edu/homes/dgleich/projects/pca_neural_nets_websites/)
[4] Hyvarinen, A.& Oja, E. Independent Component Analysis: Algorithms and applications. Neural Net.13,411-430(2000).
[5] https://medium.com/analytics-vidhya/https-medium-com-types-ofactivation-functions-in-neural-network.
[6] Recovering sound sources from embedded repetition-McDermott JH, Wrobleski D, Oxenham AJ.
[7] https://www.pnas.org/content/115/14/E3313.
[8] An overview of gradient descent optimization algorithms by Sebastian Ruder.
[9] https://www.sciencedirect.com/topics/computer-science/speakeridentification.
[10] Activation Function in Neural Network by Hamza Mahmood (https://towardsdatascience.com).
[11] M.R. Ashouri, "Isolated word recognition using high-order statistics and time delay neural network", IEEE signal processing workshop on High order statistics ,1997.
[12] https: //ieeexplore.org> voice output extraction by signal separation using deep neural network.
[13] Independent Component Analysis, a new concept?" research done by Pierre Comon,1994 (https://www.cs.purdue.edu/homes/dgleich/projects/pca_neural_nets_website/)
[14] J. Padmanabhan and M. J. Johnson Premkumar, "Machine Learning in Automatic Speech Recognition: A Survey," IETE Tech. Rev., vol. 32, no. 4, pp. 240–251, Jul. 2015.
[15] P. P. S. Pratik K. Kurzekar, Ratnadeep R. Deshmukh, Vishal B. Waghmare, "A Comparative Study of Feature Extraction Techniques for Speech Recognition System," Ijirset, vol. 3, no. 12, pp. 18006–18016, 2014.
[16] Dr. R.L.K. venkateswarlu, Dr. R. Vasantha Kumari, G. Vani Jaya Sri, International Journal of Scientific & Engineering Research Volume 2, Issue 6, June-2011.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY