DOI: http://dx.doi.org/10.26483/ijarcs.v13i1.6797

Volume 13, No. 1, January-February 2022

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

SPEECH RECOGNITION USING SOFT COMPUTING

Dr. R.K.Srivastava
CST-UP, Dr. ShakuntalaMisra National Rehabilitation University
Lucknow, India

Digesh Pandey
CST-UP, Dr. ShakuntalaMisra National Rehabilitation University
Lucknow, India

Raj Shree
CST-UP, BabasahebBhimraoAmbedkar University
Lucknow, India

Abstract: The work of speech affirmation is one of the entrancing field with respect to speech signal taking care of. Achieving accuracy and strength is a very problematic limit to various regular components. Reformist work and reviews in the speech recognition application has been gotten using Soft Computing, as one of the system to further develop the affirmation exactness's. This research paper reviews the various thoughts of Soft Computing procedure and its applications to speech signal taking care of an area. Since the possibility of speech signal is questionable, it doesn't deal with consistency at immaculate stretches. To deal with this irregularity and weaknesses, various researchers have proposed soft computing is one of the better technique to separate the speech signals. This research paper presents the composing work open related to speech recognition using Soft computing methodology.

Keywords: Speech, soft computing, accuracy, consistency, Hidden Markov Model.

I. INTRODUCTION

Speech is the key, best, solid and normal medium to impart continuously frameworks. In market because of progression in innovation numerous Speech correspondence applications based gadgets is accessible, they are less expensive and effectively accessible. Notwithstanding, undesired commotions in climate cause undesired impacts continuously speech preparing frameworks. Human interchanges and smart machines are experiences the debased presentation wherein they takes choice dependent on what it gets as a speech. Prior numerous analysts explored and created different methodologies for commotion decrease and speech upgrades. The speech upgrade is beneficial for increasing the capacity and transmission of speech data, as well as developing speech recognition-based framework execution, in which precise identification of words and sentences can provide mechanisation in the vast majority of human-machine or machine-based interfaces. By maintaining a low word blunder rate, speech upgrading can help speed up the display of voice recognition frameworks (WER). There are a variety of voice recognition frameworks available, some of which are integrated into task-specific apps. A robust Mandarin Speech Recognition framework leveraging neural networks applied to media interfaces performs better in real-world applications [5]. Speech recognition is used in a mixed media language training framework for a variety of challenges and ages. Speech recognition performs recognisable proof of speech defects and follows the patient's progress using time recurrence assessment and neural organisation methods in addition to recording the voice and breaking down the recorded spoken sign [6].

In this paper section I contains the introduction, section II contains the speech recognition system details, section III contains the details of speech recognition techniques, section IV describe the soft computing and section V provide conclusion of this paper.

II. SPEECH RECOGNITION SYSTEM

Speech recognition techniques were initially attempted in early 1952 at Bell Lab, where Davis, Biddulph, and Balashekdeveloped a disengaged digit recognition framework for a single speaker [1]. There are two types of speech ID tasks: closed set and open set. The ID of speech that already exists in the data set is included in the closed set recognisable proof; otherwise, it is an open set speech ID task. Disengaged word speech acknowledgment necessitated silence on both sides of the word, whereas ceaseless word acknowledgment makes speech difficult to perceive [2]. Apps for speech correspondence It is also used in financial structures [3, 4].

The approach of the fundamental speech acknowledgment framework [7] was proposed by Juang and B. Yegnanarayana. It consists of four basic building blocks for voice analysis: extraction, language interpretation, and message comprehension. Commotion expulsion, quiet evacuation, and end point recognition are all part of the speech examination stage. To work on the presentation of the speech acknowledgment framework, end point identification and commotion expulsion are required. Loud speech is measured along the basilar film in the internal ear, which allows for range analysis of boisterous speech.The speech analysis also maintains the suitable casing size for fragmenting speech signals for further analysis using division, sub segmental, and supra segmental examination procedures [8]. The component extraction and coding stage reduces the dimensionality of the information vector while maintaining the sign's separation force. Because the quantity of preparation and test vector required for the arrangement issue grows with the component of the given data, we need to incorporate extraction. The most often used techniques for highlight extraction are Direct

Predictive Coding (LPC) and Mel Frequency Cepstral Coefficients (MFCC). Because it is less likely to cause disruption, MFCC preferred it versus LPC.Using a neural transduction approach, the awful sign yield of speech investigation was converted to action signals on the hearable nerve. The action signal is then converted into a linguistic code within the cerebrum, and finally message comprehension is achieved.

III. SPEECH RECOGNITION TECHNIQUES

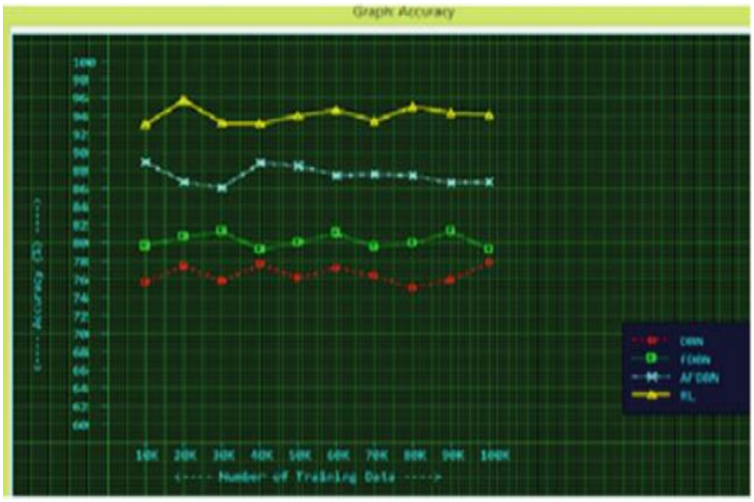
Temporal, Artificial Neural Network, and Stochastic processes are the three basic classifications for speech recognition strategies. Dynamic Time Warping (DTW) and Vector Quantization (VQ) are used for global voice recognition, while Hidden Markov Model (HMM) and Gaussian Mixture Model (GMM) are used for stochastic speech recognition, and Multilayer Perceptron is used for artificial neural network-based speech recognition (MLP). PC can use DTW to find the best match between two speech arrangements with particular constraints. The decision to be made is based on the global distance measurements between two speech designs. [9]. In DTW, there is a compromise between exactness of acknowledgment and computational productivity. Dynamic programming is used to execute enhancement measures in DTW. VQ is useful for speech coders and is commonly used in Automatic Speech Recognition (ASR). For reference models, it uses minimal codebooks. When VQ is used with DTW/HMM, capacity and computing time are reduced [10]. MLP is a neural organisation process based on back spread (BP) calculation that is used as a classifier, with hubs connected to adjacent layers by loads. The execution of MLP debases in the midst of a ruckus. Stochastic modelling is a probabilistic model arrangement with shaky data that is more appropriate for voice recognition. HMM [11] is a well-known stochastic approach that is characterised by a limited state markov model and a number of yield circulations. GMM is a mechanism for presenting text-based speech recognition. Every speaker in GMM has a free GMM model, and the yield of GMM is determined by using the most extreme probability grouping identifier.

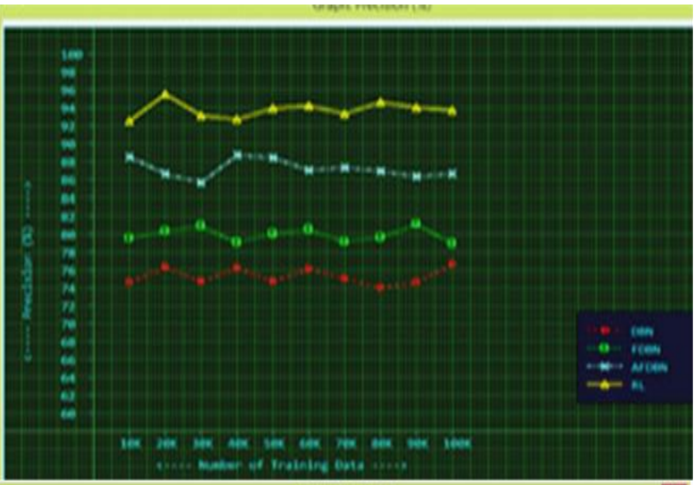
IV. SOFT COMPUTING TECHNIQUES

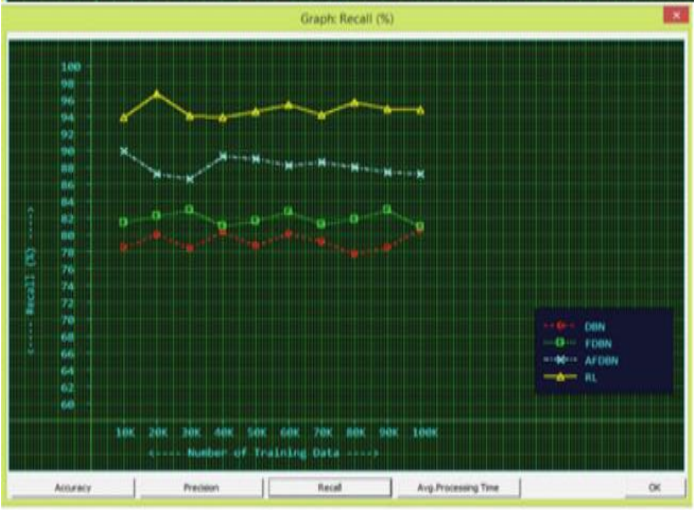
Delicate figuring is a collection of computational processes used in design disciplines to explore, show, and dissect extremely complicated problems where traditional approaches fail to provide cost-effective solutions. Neural Networks, Fuzzy Logic, and Evolutionary Computation are important components of delicate figuring (Genetic Algorithm). The Artificial Neural Network (ANN) is a data preparation paradigm inspired by the way natural sensory systems work. It is made up of a large number of extremely interconnected handling components (neurons) that work together to address certain difficulties. ANN is often used for continuous activity because ANN computations are conducted in a consistent manner. LotfiZadef devised the fluffy rationale (FL) critical thinking control framework technique. It manages ambiguous data, which is represented as fluffy sets of data. FL is used in many control framework applications because it mimics human control logic. Ga's (Hereditary Algorithms) are versatile computational approaches based on the mechanics of traditional hereditary frameworks. They used determination/generation, hybrid, and alter boundaries on a set of coded arrangements (population) [16]. The implementation and testing of HMM and ANN techniques for speech recognition on a Field Programmable Gate Array (FPGA) device was described in [17]. GA was used to prepare ANN in order to obtain a more precise and optimal arrangement. The results demonstrate that HMM has a little higher acknowledgment rate than ANN, but ANN's speech acknowledgment speed is much faster than HMM's. For the codebook plan of vector quantization, the LBG calculation is commonly used. One exciting research paper [18] offered a GA-L (GA and LBG) calculation-based approach for vector quantization in speech acknowledgment frameworks, which operates on the nature of the codebook. It's more convincing than a standard LBG calculation. The fluffy rationale acknowledgment strategy based on power conveyance example of a part of a speech continually frameworks was introduced in one research report [19]. For consistent speech preparation, example coordinating with measure is used in this paper design era. For the advanced PDA application, perspective deferral and total and versatile beamforming calculation [20] were used in the loud automobile environment.For the managed speech, performance metrics such as sign to commotion proportion and speech acknowledgment error rate were analysed in this article, and the results demonstrate that an amplifier showcase works better than a single mouthpiece framework. [21] demonstrates that a beamforming-based speech upgrading approach improves speech recognition in a multi-mouthpiece environment. The results demonstrated the discourse upgrade ability of the bar shaping strategy in multi mouthpiece organisation procedures by displaying speech acknowledgment against the channel bank boundaries; channel length and number of subbands were broken down by assessing level of acknowledgment precision, and the results demonstrated the discourse upgrade ability of the bar shaping strategy in multi mouthpiece organisation procedures.

Table 1. Comparison between different algorithms

	DBN	FDBN	AFDBN	RL
Accuracy	75%	79.4%	88.5%	94%
Precision	74%	79.4%	88.5%	93%
Recall	78%	81.4%	91%	97%
Avg. Processing Time	1145 ms	1100 ms	890 ms	820 ms







Sl. No.	Authors	Title	Purpose	Algorithm
1.	F. Beritelli	A robust voice activity detector for wireless communications using so computing	It has been demonstrated that the SNR and the threshold that minimises the total error have a linear connection.	speech compression, activity detection
2.	MohitDua , R.K.Aggarwal , Virender Kadyan ,ShelzaDua	Punjabi Automatic Speech Recognition Using HTK	For regional languages like Punjabi, an efficient, abstract, and rapid ASR system is critical.	Mel Frequency Cepstral Coefficient (MFCC), Hidden Markov Model, Dynamic Time Warp
3.	WiqasGhai, Navdeep Singh	Analysis of Automatic Speech Recognition Systems for Indo-Aryan Languages: Punjabi A Case Study	Methods' application To improve voice recognition performance, researchers are adopting Cooperative Heterogeneous ANN Architecture, Maximum Likelihood Linear Regression, Extended MFCC, and Learning Vector quantization.	Multi layer perceptron, Cooperative heterogeneous artificial neural network
4.	Milind U. Nemade , Prof. Satish K. Shah	Survey of So Computing based Speech Recognition Techniques for Speech Enhancement in Multimedia Applications	The beamforming technique is commonly utilised to improve voice recognition performance in multimedia applications.	Genetic Algorithm, Artificial Neural Network, Mel Frequency Cepstral Coefficient (MFCC)
5.	Cini Kurian	A Review on Technological Development of Automatic Speech Recognition	It is a foregone conclusion that this technology will go from machines that can somewhat duplicate human speaking skills to the building of a	HIDDEN MARKOV MODEL, Discrete Hidden Markov Model (DHMM), Semi Continuous HMM (SCHMM)

			machine that can act like an intelligent person.	
6.	Nidhi Desai , Prof.Kinn alDhameliya , Prof.Vijayendra Desai	Feature Extraction and Classification Techniques for Speech Recognition: A Review	To present a comprehensive appraisal of speech recognition and to provide some year-by-year developments to this day is a difficult and fascinating endeavour in and of itself.	Acoustic Phonetic Approach , Artificial Intelligence, Feature extraction , LPC, MFCC
7.	Gautam Krishna, Co Tran, Jianguo Yu, Ahmed H Tewfik	SPEECH RECOGNITION WITH NO SPEECH OR WITH NOISY SPEECH	created a deep learning model capable of learning EEG properties and performing speech recognition without any voice input	Electroencephalography (EEG), Speech Recognition, Distillation, Deep Learning
8.	Anupam Choudhary, Ravi Kshirsagar	Artificial Intelligence Techniques to Process Speech Recognition System	Speech recognition will be prevalent in telephone networks throughout the world over the next few years, necessitating an entirely different acoustic model. Because there is no GUI, it must be able to connect with telephony systems and manage a spoken dialogue with the user.	NLP,GUI ,IP,channel model
9.	Dr. Uma Kumari	Soft Computing Applications: A Perspective View	The goal of this work is to give a general understanding of soft computing, as well as its relevance, applications, and strengths.	Soft computing, fuzzy logic, artificial neural network, genetic algorithm

V. CONCLUSION

In this research paper, Speech is the basic, best, dependable and normal medium to impart progressively frameworks. There are such countless utilizations of speech still to be a long way from reality on account of absence of productive and solid commotion expulsion component and strategies for saving or working on the clarity for the speech signals. The purpose of this research is to look into ways for delicate registering-based speech acknowledgement procedures in interactive media apps for speech improvement. This audit showed that the beam forming method is commonly employed in mixed media applications to improve voice recognition performance. As we continue to work on the demonstration of a beam forming based speech acknowledgment framework, we may expect transformative computational calculation (GA) advances to be applied in interactive media applications. We focused on the most often used presentation estimation bounds for voice recognition.

VI. REFERENCES

[1] K. H. Davis, R. Biddulph, and S. Balashek, “Automatic Recognitionof Spoken Digits,” J. Acoust. Soc. Am., 24 (6): 637-642, 1952.

[2] M.G. Sumithra, M.S. Ramya, K. Thanuskodi, “Noise robust isolated word recognition” International Conference on Communication andComputational Intelligence (INCOCCI), pp. 362-367, 2010.

[3] A. Burstein, A. Stolzle, and R. W. Brodersen, “Using speech recognition in a personal communications system” IEEE InternationalConference on Communication, vol.3, pp.1717-1721, 1992.

[4] T. Isobe, M. Morishima, F. Yoshitani, N. Koizumi and K. Murakami, “Voice-activated home banking system and its field trial”, International Conference on Spoken Language, vol.3, pp. 1688-1691, 1996.

[5] Sheu B., Ismail M., Wang M., Tsai R., “Speech Recognition in multimedia human machine interfaces using neural networks”, Wiley-IEEE Press, pp. 463-489, 1998.

[6] V. C. Georgopoulos, “An investigation of audio-visual speech recognition as applied to multimedia speech therapy applications”, IEEE International Conference on multimedia computing and system, vol.1, pp. 481-486, 1999.

[7] L. Rabiner, B.H. Juang and B. Yegnanarayana, “Fundamentals of Speech Recognition”, Pearson Education, first edition, ISBN 978-81- 7758-560-5, 2009.

[8] H.S. Jayanna, S.R. Mahadeva, “Analysis, Feature Extraction, Modelling and Testing Techniques for Speaker recognition”, IETE Tech. Rev.,26:181-90, 2009.

[9] Bin Amin T. And Mahmood I., “Speech Recognition using Dynamic Time Warping”, Second International Conference on Advances inSpace Technologies, pp. 74-79, 2008.

[10] S. Furui, “Vector quantization based speech recognition and speaker recognition techniques”, Twenty-Fifth Asilmar Conference on Signals,Systems and Computers, vol.2, pp.954-958, 1991.

[11] A. P. Varga and R.K. Moore, “Hidden Markov Model Decomposition of Speech and Noise”, Proc. ICASSP, pp. 845-848, 1990.

[12] O. L. Frost, III, “An algorithm for linearly constrained adaptive array processing,” Proc. IEEE, vol. 60, pp. 926–935, Jan. 1972.

[13] Griffiths, L.; Jim, C., "An alternative approach to linearly constrained adaptive beamforming,",IEEE Transactions on Antennas andPropagation, vol.30, no.1, pp. 27- 34, Jan 1982.

[14] Seltzer, M.L.; Raj, B.; Stern, R.M., "Likelihood-maximizing beamforming for robust hands-free speech recognition," IEEETransactions on Speech and Audio Processing, vol.12, no.5, pp. 489- 498, Sept. 2004.

[15] W. T. Hong, “Residual Noise Removal on Beamforming for robust Hands-free Speech Recognition”, International Computer Symposium(ICS), pp. 270-273, 2010.

[16] S.N. Shivnandam, S.N. Deepa, “Principles of Soft Computing”, Wiley India Pvt Ltd, Reprint: 2010.

[17] Shing T. Pan, Ching F. Chen, Jian H. Zeng, “Speech Recognition via Hidden Markov Model and Neural Network Trained by Genetic Algorithm”, Proc. of 9th International Conference on MachineLearning and Cybernetics, Qingdao, pp. 2950-2955, 2010.

[18] Y. Yujin, Z. Qun, Z. Peihua, “Vector quantization Codebook Design Method for Speech Recognition Based on Genetic Algorithm”, Second International Conference on Information Engineering andComputer Science, pp. 1-4, 2010.

[19] Tong Zhao, Peng-Yung Woo, “Fuzzy Speech Recognition”, International Joint Conference on Neural Networks, vol. 5, pp. 2959- 2961, 1999.

[20] Stephen Oh, Vishu V., Panos P., “Hands-Free Voice Communication in an Automobile With a Microphone Array”, IEEE InternationalConference on ASSP, vol.1, pp. 281-284, 1992.

[21] Milind U. Nemade, Satish K. Shah, “Improvement in Speech Recognition Performance using Beamforming based Speech Enhancement”, International Journal of Electronics Communicationand Computer Engineering, pp. 745-751, vol.3, Issue-4, 2012.

[22] F. Beritelli, “A robust voice activity detector for wireless communications using so computing”, IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 16, NO. 9, DECEMBER 1998

[23] MohitDua , R.K.Aggarwal , VirenderKadyan ,ShelzaDua, “Punjabi Automatic Speech Recognition Using HTK”, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, July 2012

[24] WiqasGhai, Navdeep Singh, “Analysis of Automatic Speech Recognition Systems for Indo-Aryan Languages: Punjabi A Case Study”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012

[25] Milind U. Nemade, Prof. Satish K. Shah, “Survey of Soft Computing based Speech Recognition Techniques for Speech Enhancement in Multimedia Applications”, International Journal of Advanced Research in Computer and Communication Engineering, ISSN (Print) : 2319-5940 ISSN (Online) : 2278-1021, Vol. 2, Issue 5, May 2013

[26] Cini Kurian, “A Review on Technological Development of Automatic Speech Recognition”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-4 Issue-4, September 2014

[27] Nidhi Desai , Prof.KinnalDhameliya , Prof.Vijayendra Desai, “Feature Extraction and Classification Techniques for Speech Recognition: A Review”, International Journal of Emerging Technology and Advanced Engineering, (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 12, December 2013)

[28] Gautam Krishna, Co Tran, Jianguo Yu, Ahmed H Tewfik, “SPEECH RECOGNITION WITH NO SPEECH OR WITH NOISY SPEECH”, 2019, arXiv:1903.00739v1 [cs.LG] 2 Mar 2019

[29] AnupamChoudhary, Ravi Kshirsagar, “Process Speech Recognition System using Artificial Intelligence Technique”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-5, November 2012

[30] Dr. Uma Kumari, “Soft Computing Applications: A Perspective View”, 2017, Proceedings of the 2nd International Conference on Communication and Electronics Systems (ICCES 2017) IEEE Xplore Compliant - Part Number:CFP17AWO-ART, ISBN:978-1-5090-5013-0

