International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

A REVIEW PAPER ON A SYSTEM FOR SPEECH SIGNAL RECOGNITION USING VARIOUS TYPES OF ALGORITHMS

B.Kishore Babu[1], Dr.Rakesh Mutukuru [2]

[1]Research Scholar, Depatment of Ece, **Shri Venkateshwara University**, Venkateshwara Nagar, Gajraula, Uttar Pradesh-244236

[2]Professor, Depatment of Ece, **Prakasam Engineering College**,Kandukur, Andhra Pradesh-523105

ABSTRACT:

Deep learning-based machine learning models have shown significant results in speech recognition and numerous vision-related tasks. The performance of the present speech-to-text model relies upon the hyper parameters used in this research work. In this research work, it is shown that convolutional neural networks (CNNs) can model raw and tonal speech signals. Their performance is on par with existing recognition systems. Language is the most important means of communication and speech is its main medium. In human to machine interface, speech signal is transformed into analog and digital wave form which can be understood by machine. Speech technologies are vastly used and has unlimited uses. These technologies enable machines to respond correctly and reliably to human voices, and provide useful and valuable services. The paper gives an overview of the speech recognition process, its basic model, and its application, approaches and also discuss comparative study of different approaches which are used for speech recognition system. The paper also provides an overview of different techniques of speech recognition system and also shows the summarization some of the well-known methods used in various stages of speech recognition system.

Keywords: Speech recognition, LMT, DT, ANN, CNN, Data set, SR, ASR .

I INTRODUCTION

Speech signal conveys speaker information as well as linguistic information (e.g. Regional, physiological and emotional characteristics). This richness of information in speech has inspired many researches to develop the system that automatically process the speech, this speech technology has many applications [1]. Speech signal contains extremely rich information which exploits amplitude-modulated, time-modulated and frequencymodulated carriers (e.g. noise and harmonics, power, pitch, duration, resonance movements, pitch intonation) to convey information about words, speaker identity, style of speech, emotion, accent, the state of health of the speaker and expression. All these information are conveyed primarily within the traditional telephone bandwidth of 4 kHz. The speech energy above 4 kHz mostly conveys audio quality and sensation [2]. The information conveyed in speech includes the followings:

(a) Acoustic phonetic symbols. These are most elementary speech units from which larger speech units such as syllables and words are built. Some of the words have only two phones such as ‘me’, ‘you’,’ he’.

(b) Prosody. These are rhythms of speech signal carried by changes in the pitch trajectory and stress and mostly called as intonation signals. This helps to signal such information as the boundaries between segments of speech, link sub-phrases and clarify intention and remove ambiguities such as whether a sentence is a question or a statement.

(c) Gender information. Gender information is generally communicated by the pitch (related to the fundamental frequency of voiced sounds) and the size and physical characteristics of the vocal tract. Because

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

of the vocal anatomy differences, female voice usually has higher resonance frequencies and a higher pitch.

(d) Age. It is known by the effects of the size and the elasticity of the vocal cords and vocal tract, and the pitch. Children can have the pitch of voice more than 300 Hz.

(e) Accent. It broadly conveyed through:

(i) any changes in the pronunciation that will be in the form of substitution, deletion or insertion of phoneme units in the "standard" transcription of words (e.g. US Jaan pronunciation of John or Australian todie pronunciation of today) and

(ii) systematic changes in speech resonance frequencies (formants), emphasis, stress and pitch intonation, duration.

(f) Speaker's identity is known by the physical characteristics of a person's vocal folds, vocal tract, pitch intonations and stylistics.

(g) Emotion and health is known by changes in: vibrations of vocal fold, vocal tract resonance, duration and stress and by the dynamics of pitch and vocal tract spectrum.

Definition of speech recognition: Speech Recognition (is also known as Automatic Speech Recognition (ASR), or computer speech recognition) is the process of converting a speech signal to a sequence of words, by means of an algorithm implemented as a computer program.

Main objective of research:

Research in speech processing and communication for the most part, was motivated by people s desire to build mechanical models to emulate human verbal communication capabilities. Speech is the most natural form of human communication and speech processing has been one of the most exciting areas of the signal processing. Speech recognition technology has made it possible for computer to follow human voice commands and understand human languages. The main goal of speech recognition area is to develop techniques and systems for speech input to machine. Speech is the primary means of communication between humans. For reasons ranging from technological curiosity about the mechanisms for mechanical realization of human speech capabilities to desire to automate simple tasks which necessitates human machine interactions and research in automatic speech recognition by machines has attracted a great deal of attention for sixty years. Based on major advances in statistical modeling of speech, automatic speech recognition systems today find widespread application in tasks that require human machine interface, such as automatic call processing in telephone networks, and query based information systems that provide updated travel information, stock price quotations, weather reports, Data entry, voice dictation, access to information: travel, banking, Commands, Avoinics, Automobile portal, speech transcription, Handicapped people (blind people) supermarket, railway reservations etc. Speech recognition technology was increasingly used within telephone networks to automate as well as to enhance the operator services. This report reviews major highlights during the last six decades in the research and development of automatic speech recognition, so as to provide a technological perspective. Although many technological progresses have been made, still there remains many research issues that need to be tackled.

Types of Speech Recognition:

Speech recognition systems can be separated in several different classes by describing what types of utterances they have the ability to recognize. These classes are classified as the following:

Isolated Words: Isolated word recognizers usually require each utterance to have quiet (lack of an audio signal) on both sides of the sample window. It accepts single words or single utterance at a time. These systems have "Listen/Not-Listen" states, where they require the speaker to wait between utterances (usually doing processing during the pauses). Isolated Utterance might be a better name for this class.

Connected Words: Connected word systems (or more correctly 'connected utterances') are similar to isolated

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

words, but allows separate utterances to be 'run-together' with a minimal pause between them.

Continuous Speech: Continuous speech recognizers allow users to speak almost naturally, while the computer determines the content. (Basically, it's computer dictation). Recognizers with continuous speech capabilities are some of the most difficult to create because they utilize special methods to determine utterance boundaries.

Spontaneous Speech: At a basic level, it can be thought of as speech that is natural sounding and not rehearsed. An ASR system with spontaneous speech ability should be able to handle a variety of natural speech features such as words being run together, "ums" and "ahs", and even slight stutters.

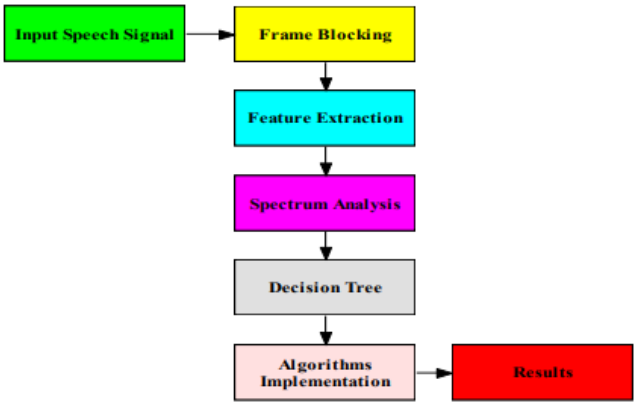

Fig.1. Basic model of speech recognition system.

II SURVEY OF RESEARCH

[1] Applications in Speech Recognition System by Rashmi C R explain about Speech is one of the natural ways for humans to communicate. Human Voice is a unique characteristic for any individual. A valuable biometric tool can be designed based on the ability to recognize a person by his/her voice and this biometric tool has enormous commercial as well as academic potential. It can be commercially used for ensuring secure access to any system. This paper delivers an overview of different algorithms that can be used in applications of speech recognition based on the advantages & disadvantages. It also helps in choosing the better algorithm based on the comparison done.

[2] Speech Recognition by Machine: A Review by M .A.Anusuya explain about in this paper presents a brief survey on Automatic Speech Recognition and discusses the major themes and advances made in the past 60 years of research, so as to provide a technological perspective and an appreciation of the fundamental progress that has been accomplished in this important area of speech communication. After years of research and development the accuracy of automatic speech recognition remains one of the important research challenges (eg., variations of the context, speakers, and environment).The design of Speech Recognition system requires careful attentions to the following issues: Definition of various types of speech classes, speech representation, feature extraction techniques, speech classifiers, database and performance evaluation. The problems that are existing in ASR and the various techniques to solve these problems constructed by various research workers have been presented in a chronological order. Hence authors hope that this work shall be a contribution in the area of speech recognition. The objective of this review paper is to summarize and compare some of the well known methods used in various stages of speech recognition system and identify research topic and applications which are at the forefront of this exciting and challenging field.

[3] Introduction to Various Algorithms of Speech Recognition: Hidden Markov Model, Dynamic Time Warping and Artificial Neural Networks by Pahini A. Trivedi explain about speech recognition is used widely in many applications. In computer science and electrical engineering, speech recognition (SR) is the translation of spoken words into text. It is also known as "automatic speech recognition" (ASR), "computer speech recognition", or just "speech to text" (STT). A hidden Markov model (HMM) is a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be presented as the simplest dynamic Bayesian network. Dynamic time warping (DTW) is a well-known technique to find an

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

optimal alignment between two given (time-dependent) sequences under certain restrictions intuitively; the sequences are warped in a nonlinear fashion to match each other. ANN is non-linear data driven self-adaptive approach. It can identify and learn co-related patterns between input dataset and corresponding target values. After training ANN can be used to predict the outcome of new independent input data.

[4] Speaker Accent Recognition Using Machine Learning Algorithms author by Ahmet Aytu Ayranc explain in this paper Speaker recognition is a system that recognizes the speaker from the recorded voice signal. Speech and speaker recognition are important for many areas like online banking, telephone shopping, and security applications. In order to analyze and verify speech and speaker, Machine Learning (ML) algorithms can be used. With enough data it is possible to train a program to identify speech and speaker identity. In this paper, several ML algorithms used to identify speaker accents. Data set includes 329 speakers with 6 different accents and English and these words have been converted to metric representation using Mel-Frequency Cepstral Coefficients (MFCCs). MFCC is the most used technique in speech recognition because of the high performance of feature extraction performance and this data set utilizes MFCC to convert speech to data. In this study, 7 classification type ML algorithms used, including Multilayer Perceptron (MLP), Random Forest (RF), Decision Tree (DT), Radial Basis Function (RBF), k-Nearest Neighbor (k-NN), Naive Bayes (NB) and Logistic Model Tree (LMT) methods used for Speaker Accent Recognition data set on UCI ML Repository. Performance metrics, compared using accuracy, Kappa Statistics, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), have been acquired and compared for each algorithm.

[5] Speech Recognition System – A Review author by Shaikh Naziya S.1* , R.R. Deshmukh2 explain Language is the most important means of communication and speech is its main medium. In human to machine interface, speech signal is transformed into analog and digital wave form which can be understood by machine. Speech technologies are vastly used and has unlimited uses. These technologies enable machines to respond correctly and reliably to human voices, and provide useful and valuable services. The paper gives an overview of the speech recognition process, its basic model, and its application, approaches and also discuss comparative study of different approaches which are used for speech recognition system. The paper also provides an overview of different techniques of speech recognition system and also shows the summarization some of the well-known methods used in various stages of speech recognition system.

[6] Detection and Analysis of Emotion From Speech Signals author by Assel Davletcharova *et al,* analyzed and detect the emotion from speech signals. Perceiving emotion from speech has turned out to be one the dynamic research topics in speech processing and in applications in the view of interaction between human and system. This paper directs a study of emotion recognition from human speech. The emotion considered for the analyses such as happiness, sadness, anger or neutral. These are studied using the emotion classification performed on a custom dataset. It performed for various classifiers. One of the principle highlight attribute considered in the arranged dataset was the distance from end to end acquired from the graphical representation of the speech signals. Subsequently performing datasets shaped from 30 unique subjects, it was found that for improving accuracy, one ought to consider the information gathered from one individual instead of considering the information from a gathering of individuals.

[7] Use of SVM Classifier & MFCC in Speech Emotion Recognition System by Bhoomika Panda *et al* uses the SVM classifier and MFCC in the emotion recognition system for speech signal. Speech Emotion Recognition (SER) is a recent research topic in the field of Human Computer Interaction (HCI) with extensive variety of utilizations. The speech elements, like Mel Frequency cepstrum coefficients (MFCC)

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

extract the expression of emotion. The Support Vector Machine (SVM) is utilized as classifier to order extraordinary enthusiastic states, like fear, anger, sadness, from a database of emotional speech gathered from different enthusiastic show sound tracks. The SVM is utilized for grouping of feelings. It gives 93.75% characterization exactness for Gender free case 94.73% for male and 100% for female speech.

[8] Speech-Based Emotion Recognition: Feature Selection by Self-Adaptive Multi-Criteria Genetic Algorithm author by Maxim Sidorov describe speech based emotion recognition by extracting the feature selection with the help of self adaptive genetic algorithm. The emotion recognition has various applications in Interactive Voice Response frameworks, call centers etc. While utilizing existing capabilities and strategies for automatic emotion recognition has as of now accomplished sensible outcomes, there is still a considerable measure to accomplish for development. In the research most of the features are extracted using this genetic algorithm for performing speech based emotion recognition is still an open question. Thus it improves the performance for the applied databases.

[9] A Real Time Classifier for Emotion and Stress Recognition in a Vehicle Driver author by M. Paschero proposed a real time classifier in a vehicle driver for recognizing their stress and emotion. As of late there is an incredible enthusiasm for artificial system ready to realize and perceive human emotions. In this paper an Emotion Recognition System in light of classical neural systems and neuro-fuzzy classifiers is proposed. The emotion recognition is performed continuously beginning from a video stream obtained by a typical webcam checking the face of user. Neuro-fuzzy classifiers, in examined with Multi Layer Perceptron prepared by EBP algorithm, demonstrate short preparing circumstances, permitting applications with simple and computerized set up strategies, to be utilized as a part of an extensive variety of utilizations, from stimulation to wellbeing. The algorithm yields extremely intriguing exhibitions and can be received to perceive emotions and in addition conceivable neurotic states of the person to be checked.

[10] Stress and Emotion Recognition Using Log-Gabor Filter Analysis of Speech Spectrograms author by Ling He use the Log-Gabor filter analysis for stress and emotion recognition with speech magnitude spectrograms. In this shows a new techniques that performs feature extraction from discourse size spectrograms. Two of the introduced approaches have been discovered especially effective during the process stress and emotion characterization. In the principal approach, the spectrograms are sub-isolated into ERB frequency bands and the average energy for every band is figured. In the second approach, the spectrograms are gone through a bank of 12 log-Gabor channels and the output are averaged at the midpoint of and went through an optimal feature extraction based on the criteria of mutual information.

III RECOMMENDED SYSTEM

Speech recognition technologies allow computers to interpret human speech into text by computers. Some SR systems use training where an individual speaker records text into the system. The system analyzes the person's particular voice and employs it to fine-tune the recognition of that person's speech, resulting in enhanced accuracy. Systems that do not use training are termed as speaker independent systems. Systems that use training are called speaker dependent. Recording through microphone as input device is considered as an acoustic Signal. Signal is then translated into a text. The weakness especially comes from the level of accuracy in speech recognition.

We have planned to propose an FNN for recognition of speech. When an input pattern is provided, the network first fuzzifies this pattern and then co-relates the speech to all of the learned ones. Fuzzy neural network (FNN) combined by neural network and fuzzy system, can mimic the human brain logic thinking. A voice analysis is done after taking an

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

input through microphone from a user. The design of the system involves manipulation of the input audio signal. The extracted features are adapted by means of an Optimization method employing the Motley Bat Algorithm.

Objectives:

- To propose an FNN for recognition of speech
- To design the system that involves manipulation of the input audio signal
- To design Motley BAT optimization algorithm, this is the combination of Levy flight and BAT algorithm
- To improve the classification accuracy using FNN

The outcome of the work will be the accomplishment of accuracy for Speech recognition using Fuzzy Neural Network. Under various circumstances the output of the system will be compared to show the system effectiveness.

Methodology:

This approach is designed for complicated tasks but it is not as efficient as HMM in the case of large vocabularies. Phoneme recognition is the general approach of neural networks. In this approach the technique of intelligence, analyzing and visualizing of speech signal is done to measure phonetic features. The network includes a huge number of neurons. Each neuron computers nonlinear weight of inputs and broadcast result to the outgoing units, training sets are used for assigning pattern of values to input and output neurons, training set determines the weight of strength of each pattern.

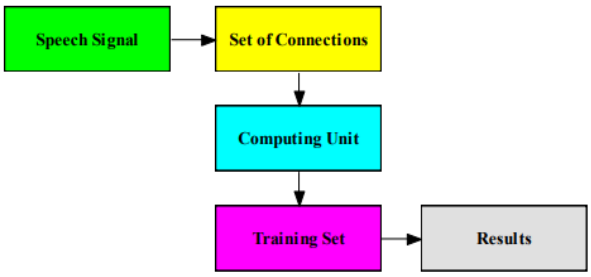

Fig.2. Expected model for speech recognition system.

CONCLUSION

Speech Recognition System is growing day by day and has unlimited applications. The study has shown the overview of the speech recognition process, its basic model, and applications. In this study total seven different approaches which are widely used for SRS have been discussed and after comparative study of these approaches it is concluded that Hidden markov method is best suitable approach for a SRS because it efficient, robust, and reduces time and complexity. Speech recognition is one of the most integrating areas of machine intelligence, since; humans do a daily activity of speech recognition. Speech recognition has attracted scientists as an important discipline and has created a technological impact on society and is expected to flourish further in this area of human machine interaction. We hope this paper brings about understanding and inspiration amongst the research communities of ASR.

REFERENCES

[1] De Silva, L.C. "Audiovisual Emotion Recognition". In proceeding of international International Conference on Systems, Man and Cybernetics, 2004.

[2] Picard, Rosalind W. "Automating the Recognition Of Stress And Emotion: From Lab To Real-World Impact". Journal of IEEE Multi Media Vol.23, No.3, Pp: 3-7, 2016.

[3] Laxmi Narayana, M and Sunil Kumar Kopparapu. "On The Use of Stress Information in

Speech for Speaker Recognition". In proceeding conference of TENCON ,2009.

[4] Bou-Ghazale, S.E. and J.H.L. Hansen. "A Comparative Study of Traditional And Newly Proposed Features For Recognition Of Speech Under Stress". Journal of IEEE Transactions on Speech and Audio Processing, Vol. 8, No.4, and Pp: 429-442, 2000.

[5] Y. H. Gulhane, S. A. Ladhake "A Short Survey on Methodology for Stress Recognition", Journal of Emerging Research in Management &Technology, Vol.5, No.11, 2016.

[6] Davletcharova, Assel et al. "Detection and Analysis of Emotion From Speech Signals". Journal of Procedia Computer Science Vol. 58 Pp: 91-96, 2015.

[7] Bhoomika Panda et al" Use of SVM Classifier & MFCC in Speech Emotion Recognition System", Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, No.3, 2012.

[8] Maxim Sidorov Speech-Based Emotion Recognition: Feature Selection by Self-Adaptive Multi-Criteria Genetic Algorithm", in proceedings of IREC conference, 2014.

[9] Paschero, M. et al. "A Real Time Classifier for Emotion and Stress Recognition in a Vehicle Driver". In proceeding conference of IEEE International Symposium on Industrial Electronics, 2012.

[10] He, Ling et al. "Stress and Emotion Recognition Using Log-Gabor Filter Analysis of Speech Spectrograms". 2009 In proceeding of International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009.

[11]. GIN-DER WU AND YING LEI (2008), "A register array based low power FFT processor for speech recognition. "JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 3 (2008).

[12]. Juang, B. H., & Rabiner, L. R. (2005). Automatic speech recognition–A brief history of the technology development. Encyclopedia of Language and Linguistics.

[13]. King, S., Frankel, J., Livescu, K., McDermott, E., Richmond, K., & Wester, M. (2007). Speech production knowledge in automatic speech recognition. The Journal of the Acoustical Society of America, 121(2), 723-742.

[14]. Klevans, R. L., & Rodman, R. D. (1997). Voice recognition. Artech House, Inc.

[15]. Maheswari, N. U., Kabilan, A. P., & Venkatesh, R. (2010). A hybrid model of neural network approach for speaker independent word recognition. International Journal of Computer Theory and Engineering, 2(6), 912.

[16]. Moore, R. K. (1994, September). Twenty things we still don't know about speech. In Proc. CRIM/FORWISS Workshop on Progress and Prospects of speech Research and Technology.

[17]. Morales, N., Hansen, J. H., & Toledano, D. T. (2005, March). MFCC Compensation for Improved Recognition of Filtered and Band-Limited Speech. In ICASSP (1) (pp. 521-524).

[18]. Reddy, D. R. (1966). Approach to computer speech recognition by direct analysis of the speech wave. The Journal of the Acoustical Society of America, 40(5), 1273-1273.

[19]. Shaughnessy, D. O. (2003). Interacting with computers by voice: automatic speech recognition and synthesis. Proceedings of the IEEE, 91(9), 1272-1305.

[20]. Weintraub, M., Murveit, H., Cohen, M., Price, P., Bernstein, J., Baldwin, G., & Bell, D. (1989, May). Linguistic constraints in hidden Markov model based speech recognition. In Acoustics, Speech, and Signal Processing, 1989.

ICASSP-89., 1989 International Conference on (pp. 699-702). IEEE.