International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

An Appraisal on Speech and Emotion Recognition Technologies based on Machine Learning

C Andy Jason, Sandeep Kumar

Abstract**: *In earlier days, people used speech as a means of communication or the way a listener is conveyed by voice or expression. But the idea of machine learning and various methods are necessary for the recognition of speech in the matter of interaction with machines. With a voice as a bio-metric through use and significance, speech has become an important part of speech development. In this article, we attempted to explain a variety of speech and emotion recognition techniques and comparisons between several methods based on existing algorithms and mostly speech-based methods. We have listed and distinguished speaking technologies that are focused on specifications, databases, classification, feature extraction, enhancement, segmentation and process of Speech Emotion recognition in this paper.***

Keywords: Speech Emotion Recognition, Speech Processing, Biometric, Machine Learning, MLP.

I. INTRODUCTION

The vocal cord generates sounds via disc-platelets and a series of vibrations, which give the production a speech signal, as air is breathed from the lungs. Speech processing in many specialized workstations incorporating software and telephony will make use of the growing overlap between information processing and conventional transport of information [1-2]. More recently, the interest in automatically collecting vast quantities of voice data was growing to establish not only what was being said, but also how and by whom [3]. The present device becomes a personal workstation mobile terminal with added speech input and output capabilities that allow giving access around the world [4-6]. In another case the time spent on the telephone mark on unproductive play can be cut down by apps like voice messaging, remote access to the e-mail using text to speech synthesis and recognizing the speech. But when it comes to speech emotion recognition there are many classifiers and the latest methods which are used to improve accuracy are briefly mentioned in this paper.

Manuscript published on January 30, 2020.
***** Correspondence Author

Sandeep Kumar*, Professor, Sreyas Institute of Engineering and Technology, Hyderabad, India. Email: drsandeep@sreyas.ac.in

C. Andy Jason, M.Tech Student, Sreyas Institute of Engineering and Technology, Hyderabad, India. Email: jasonandy100496@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

We use various speech databases to build problem-solving strategies and techniques specific to languages [7-10]. In those, few are briefly mentioned in the literature review to learn more about the techniques. Digital speech processing aims at using digital computing technology to process speech signals for better understanding and increased efficiency of interaction and productive connection with speech activities. Let's look at the history of the various technologies in recent decades for speech recognition [11-14].

With the increasing use of voice as a biometric, voice has become a key factor in the generation of speech. A conversation is being prepared to become an evolving technology that enables individuals to communicate with machines. From this point forward, it was possible to develop speech recognition software [12-17]. In 1922, the major production began. The name of the device is called Rex which is a celluloid dog that jumped when 500 Hz of acoustic energy was released out with the help of spring. Because the first vowel formant is usually 500 Hz in "Rex," the dog seemed to come when it had been named. A shoebox is a revolutionary gadget that is written in 16 letters, including a decadence of' 0 to 9 which have been verbally expressed and addressed. Shoebox which was developed by IBM in the year 1962 taught the calculator to calculate and print responses to simple problems with counting. HEARSAY-I and DRAGON developed the HARPY model in CMU during 1976 following a basic analysis [18-21]. This includes all learning stages: auditory, phonemic, lexical, syntactic and semantic. Hidden Markov (HMMs, 1980) models are recognized for their application in enhancement learning and the recognition of temporal patterns like expressions, handwriting, gesture recognition, speech markings, follow-up music performance, partial discharges as well as bioinformatics [22-23]. A deep neural network that was invented before 2010 was a multi-level neural network with a certain degree of predictability [24-28]. Deep neural networks use advanced scientific evidence for complex processing of information. This deep neural network is primarily used in current language recognition software in present speech-related technologies [29-30]. To detect and process the audio signals, several techniques and inventions are used to plan the necessary speech extracted from the discrete audio signal [31-36].

II. TRADITIONAL TECHNIQUES FOR SPEECH EMOTION RECOGNITION

A. Various Classifiers Based on Enhancement

Enhancement of speech is used to improve speech performance through different measurements.

The goal is to improve the understandability and overall sensory performance of distorted speech signals by using processing techniques related to speech algorithms [32].

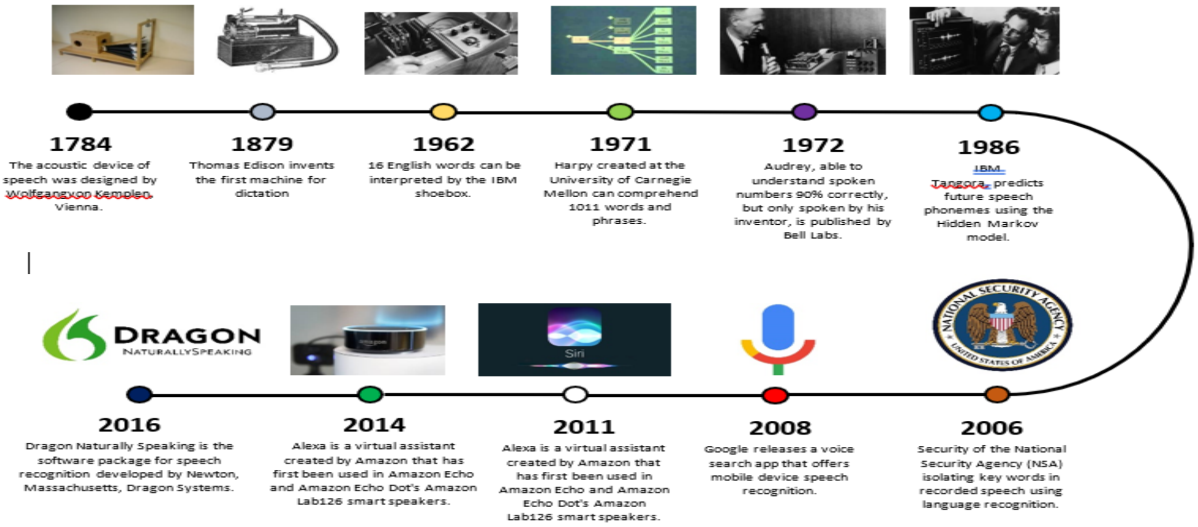
1784
The acoustic device of speech was designed by Wolfgangvon Kemplen, Vienna.
1879
Thomas Edison invents the first machine for dictation
1962
16 English words can be interpreted by the IBM shoebox.
1971
Harpy created at the University of Carnegie Mellon can comprehend 1011 words and phrases.
1972
Audrey, able to understand spoken numbers 90% correctly, but only spoken by his inventor, is published by Bell Labs.
1986
IBM Tangora, predicts future speech phonemes using the Hidden Markov model.
DRAGON
NATURALLYSPEAKING
Siri
2016
Dragon Naturally Speaking is the software package for speech recognition developed by Newton, Massachusetts, Dragon Systems.
2014
Alexa is a virtual assistant created by Amazon that has first been used in Amazon Echo and Amazon Echo Dot's Amazon Lab126 smart speakers.
2011
Alexa is a virtual assistant created by Amazon that has first been used in Amazon Echo and Amazon Echo Dot's Amazon Lab126 smart speakers.
2008
Google releases a voice search app that offers mobile device speech recognition.
2006
Security of the National Security Agency (NSA) isolating key words in recorded speech using language recognition.

Fig 1: List of Development in Technologies used for Speech Recognition [1]

The main areas of speech enhancements include noise reduction and other applications, such as smartphones utilities, teleconference, recognition of speech and audio equipment. Three simple classes can group speech enhancement algorithms for noise reduction: spectral restoration, methods based on the model and filtering techniques.

1 Filtering Techniques

- *Spectral Subtraction Method:* Spectral subtraction is a technique used to recover signal energy or amplitude in additional noise by extracting an approximation of the normal noise range from the noisy signal spectrum [33-37], [65].
- *Wiener Filtering for enhancement:* It is a filter that uses noise range, stationary signal, known and additive noise to approximate a desired or target the random process and observed noise stream by filtering a linear time-invariant (LTI) [38-40], [66].
- *Signal subspace approach:* The Signal subspace approach then generates a certain quantity of noise filtered by linear combinations between the first few basic, most energy-efficient vectors [41-42]. Linear theoretical approaches to dimension reduction and noise reduction are signal subspace methods. All methods recently have received considerable attention and focus on learning vocabulary, speech modeling and the recognition of language [67-68].

2 Spectral Restoration

- Spectral restoration is a process by subtracting from the noise spectrum an approximation of the normal noise spectrum for restoring the signal strength or frequency range found in the additional noise. Since the signal is absent, the distribution of noise is calculated and adjusted and only the sound is present [43-45], [67].
- *STSAE-MMSE method based on restoration technique* This method uses a minimum mean-square error estimator which is developed and then compared to other commonly used filtering methods and the spectral subtraction algorithm [68-69].
- *Speech-Model-Based:* Those models depend on the Speech Enhancement Technique application and speech algorithms and many efficient technologies and algorithms used to boost the speech [46-47].

B. Various classifiers Based on Classification

Segmentation of speech [48] is the method of defining the endpoints in spoken natural languages between words, syllables, or phonemes. The term applies to physical human processes as well as to natural speech processing. The algorithms and methods which are based on the Segmentation of Speech are:

- Wavelet Method: This method [49-51] is used to identify the beginnings and ends of the phonemes which are based on the analysis of DWT [52] and, with the aid of spectral analysis, an extremely effective technique has occurred in the form result [53].
- Artificial Neural Networks: An ANN is based on an artificial neuron sequence of linked units or knots, which loosely from the neurons of a biological brain [54]. The signal can be transmitted from each connection by another neuron, like synapses within a neural network of Brain [55].
- Blocking Black Area Method: This technique is employed for blocking areas of voice so that speech-speaking parts from silence or voiceless parts of voice can be easily distinguished. For continuous speech, the edges of the block are used as word limits. The crucial function of the segmentation of speech is to define the boundaries of the speech system [56].
- Short Term Energy: E. A. Kaur, E. T. Singh used the speaking method and Short-Term Energy to segment speech into syllables. MATLAB 7.8 implemented the technique. The proposed

approach was implemented and analyzed for different Punjabi speech signals [50-57].

- Hybrid Speech Segmentation Algorithm: In the Segmentation of Hybrid Speech algorithm, the features were extracted to determine the limits of the term using a basic threshold methodology.
- This characterizes the segmentation to break down continuous speech into a series of terms or sub-words. The segmentation method was 98.33% successful and the error rate was 1.67%. [58].
- *Word Chopper Technique:* P. Singh, N. Sharma proposed the approach of interrupting speech into syllables. using Word Chopper. They proposed a new method involving three steps of characteristics removal, matching rules and segmentation. [59].
- *Hidden Markov Model:* Hidden Markov System is a Markov statistical model in which the simulation system is believed to be a Markov process with non-compliant states. It is possible to describe the hidden Markov model as the simplest dynamic Bayesian network [60]. These are a few recent techniques and algorithms used for the segmentation of speech.

C. Various classifiers Based on Feature Extraction

The main part of the speech recognition process is the retrieval of information. It is known as the core of the system. The aim of this is to extract certain features that help the machine recognize the speaker from the input speech (signal). Few algorithms are used to remove features based on technology for speech recognition. They are.

- *Linear Predictive Coding:* LPC is a device used to process expression. LPC is based on an assumption: we must predict the nth sample in a sequence of speech samples that can be interpreted by summing up the previous samples of the target signal. An inverse filter should be generated in such a way that it corresponds to the forming regions of the speech samples. The use of these filters in the samples is, therefore, the LPC process. [70].
- *MFCC Method:* It is known variations of the critical range based and the main purpose of this processor is to copy human ear behavior with below 1000 Hz frequencies [71].
- *RASTA Filtering:* It is a method which is long for relative spectral used when filmed in a noisy environment to improve speech. The time coefficients and trajectories of signal speech representations are filtered bandpasses in the algorithm [72-73].
- *Probabilistic Linear Discriminate Analysis (PLDA):* This technique is a linear probabilistic analysis extension. This technique was initially used to recognize the face but is now used to recognize speech [74]. These are just a few new and latest technologies and algorithms that are used in speech-based function extraction using these analysis methods.

D. MLP Classifier

A deep, artificial neural network is a multilayer perceptron (MLP). It consists of several perceptrons. It is composed of an input layer for receiving the signal, an output layer for deciding or predestining the data, and an arbitrary number of hidden layers, between them, which represent the MLP's true machine engine. MLPs with a hidden layer can approximate any permanent feature.

$$y = \varphi\left(\sum_{i=1}^{n} \omega_i x_i + b\right) = \varphi(W^T X + b)$$

Multi-layered perceptrons are often used to monitor learning problems: they learn about the correlations (or dependencies) between those inserts and outputs and train them on a number of input-output pairs. Training consists of changing the model's parameters, weights and biases in order to minimize errors. Backpropagation is used to adjust the weight and the bias relative to the error and the error can be measured in a range of ways, including by root average squared error. MLP trains using Stochastic Gradient Descent, Adam, or L-BFGS. Stochastic Gradient Descent (SGD) updates parameters using the gradient of the loss function with respect to a parameter that needs adaptation, i.e.

$$\omega \rightarrow \omega - \eta = \alpha \frac{\partial R(\omega)}{\partial \omega} + \frac{\partial Loss}{\partial \omega}$$

where η is the learning rate which controls the step-size in the parameter space search. Loss is the loss function used for the network.

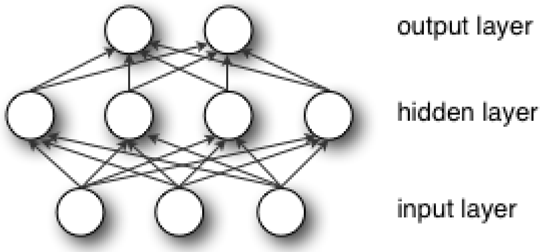

Fig 2: Outline of MLP Classifier

Networks like MLPs are like ping-pong or tennis. Two motions, constant back and forth are involved in the first instance [47], [53-56]. This is a kind of accelerated science since every guess is a test of what we think we know and every answer is feedback so that we know how wrong we are. It is a test of what we think we know.

The signal stream from the input layer to the output layer is transferred from the hidden layers and the decision is taken against the ground truth labels. In the reverse pass, partial derivatives of the error function w.r.t. are reproduced via MLP using the rear spread and the chain rule of calculus. This defining act gives us a gradient or error environment where the parameters can be changed as the MLP moves one step closer to a minimal error. This can be achieved with any optimization algorithm dependent on gradients, like the stochastic gradient descent [61-64].

- *MLP Classifier vs Other Classification Algorithms:* MLP Classifier stands for Multi-layer Perceptron classifier which in the name itself connects to a Neural Network. Unlike other classification algorithms such as Support Vectors or Naive Bayes Classifier, MLP Classifier relies on an underlying Neural Network to perform the task of classification [69-70]. Let's now see how speech is produced and recognized using speech technologies and algorithms using the diagram below.

III. METHODOLOGY

The Production and generation of Speech are recognized in four major steps.

Activities that are related to speech, procurement, processing and generation of output and application of based on the function of the Speech extracted. Here, as vocal cords are used to produce speech from the mouth, sensors and recording techniques can detect it. The voice will then be processed using expression techniques and processes. Speech synthesis then takes place to rearrange the voice bits to a certain frequency to produce a coherent word that is further used by the product based on the use of speech technology.

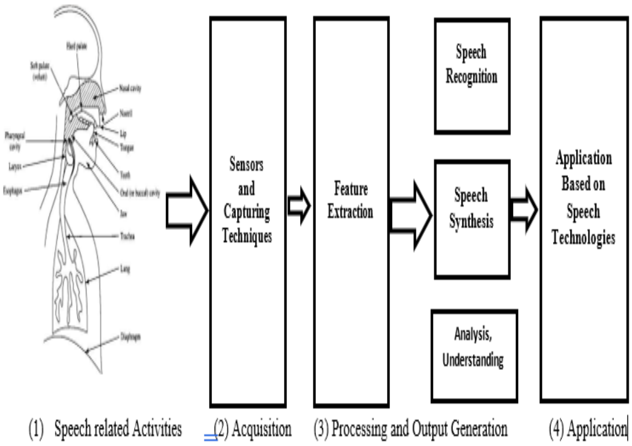
Hard palate
Soft palate (velum)
Nasal cavity
Nostril
Lip
Tongue
Teeth
Oral (or buccal) cavity
Jaw
Trachea
Lung
Diaphragm
Pharyngeal cavity
Larynx
Esophagus
Sensors and Capturing Techniques
Feature Extraction
Speech Recognition
Speech Synthesis
Analysis, Understanding
Application Based on Speech Technologies
(1) Speech related Activities
(2) Acquisition
(3) Processing and Output Generation
(4) Application

Fig 3: Speech generation and Recognition based on the application of Speech Technologies [46]

But when it comes to Speech Emotion recognition, the process is explained in five main steps and these steps are involved in each and every application related to SER which is based on machine learning techniques [75]. They are classified as Speech input, Feature Extraction, Feature Selection, Classifier and Recognition of Emotion in Speech [76]. They are

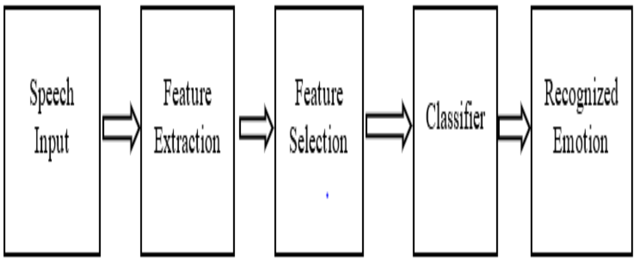
Speech Input
Feature Extraction
Feature Selection
Classifier
Recognized Emotion

Fig 4: Process of Recognizing Emotion from Speech

A typical set of emotion contains 300 Emotional States. While the primary emotions are Anger, Disgust, Fear, Joy, Sadness, and Surprise. In those, Anger, Joy, Sadness and fear are primary[77]. These emotions are differentiated based on the corresponding changes that occur in speech rate, pitch, energy and spectrum one of the main speech features that indicate emotions is energy and study of energy [78]. It depends on the short-term energy and short-term average amplitude. The speech input is taken based on the required format and then feature extraction is done by MFCC and LPCC methods and selection process based on the Artificial Neural Network (ANN) which are Machine Learning Techniques. After that, there are various types of classifiers are used and finally, emotion is recognized as shown in the above block diagram.

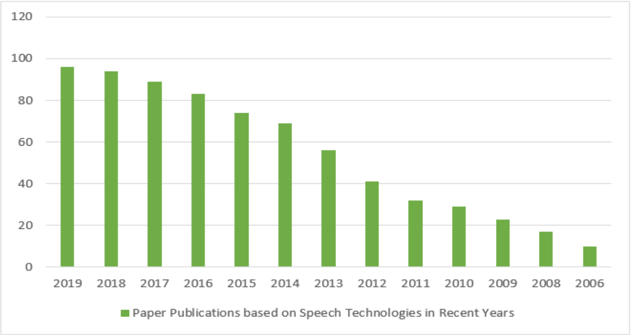

Fig 5: Recent statistics of Papers Published based on Speech Technologies [3-12]

There are different phases of speech processing that generate beautiful difficulties when understanding speech. To date, several researchers have built several speech databases to evaluate the quality of speech applications, but they lack the efficiency to create successful speech applications due to less speaking data available on them which are in terms of time periods. In recent years, interest in speech technology has gradually increased in parallel with increased accuracy, as shown in the following graphs.

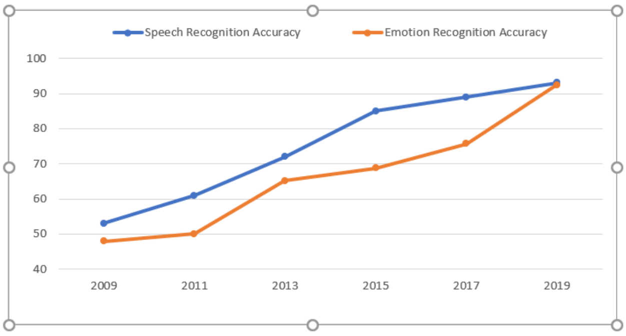

Fig 6: Growth in Speech and Emotion Recognizing Accuracy [2]

For several decades, Emotion Recognition using speech is the main challenging and essential topic in the field of human and computer communication [79-80]. To use this, the task of understanding the feelings of a speaker from the speech they are captured is all essential. All right, several efforts have been made to recognize emotions using several classification strategies for machine learning [81-82]. The benefits of using some emotion recognition information are recent and most useful for recognizing emotional expression in a speech in a number of organizations. The current solutions to this issue include supporting vector machines, Markov models and neural networks. As we know that vector machines offer better results and less effort, hidden Markov and neural network models cannot be designed and trained. Now let's see-through a few kinds of literature review the classifications of different speech-based technologies [83].

IV. RELATED WORK

A. Classification of Speech Processing:

Let us see how the classification of Speech Processing is done base on each and every aspect and application by observing the following diagram.

Speech processing is developing and increasing by data processing and providing traditional transportation for many workplaces that integrate application-based computing and telephony through different platforms. The present device becomes a terminal for the personal remote workstation with added speech input and output capabilities, allowing access to its functionality from anywhere depending on its result and outcome [84]. The required speech is also minimized by the advent of effective and spectral algorithms, very large circuits with integrated and processors of digital signals and detection of the audio signal based on these efficient algorithms and methods. Speech recognition is typically used to translate the spoken word to a specific speech message response, while speech verification is used to check the voice features of the clients. The aim of speech recognition systems is simply to understand the speaker's spoken word and develop the speaker's identity.

A common instance of speech recognition is that of an automated call center that asks a client to press the number one on their phone keyboard or say the word one [85]. In this situation, the machine does not confirm the identity of the person who says the word one it simply checks that the word one has been spoken instead of another. The need for Speech processing is to apply virtual computing techniques to process the voice signal for increased understanding, connectivity, and speech-related which are based on a quality product. In speech technologies related to processing, text-to-speech synthesis, spoken language dialogue systems, digital voice coding and ASR’s are used. Knowledge (such as speaking identity, gender, vocabulary, and speech recognition) can also be derived from speech. Speech Processing applications in real-time are talking to robots, controlling digital devices, supporting visually impaired and hearing-impaired technology, and hands-free technology based on speech [86].

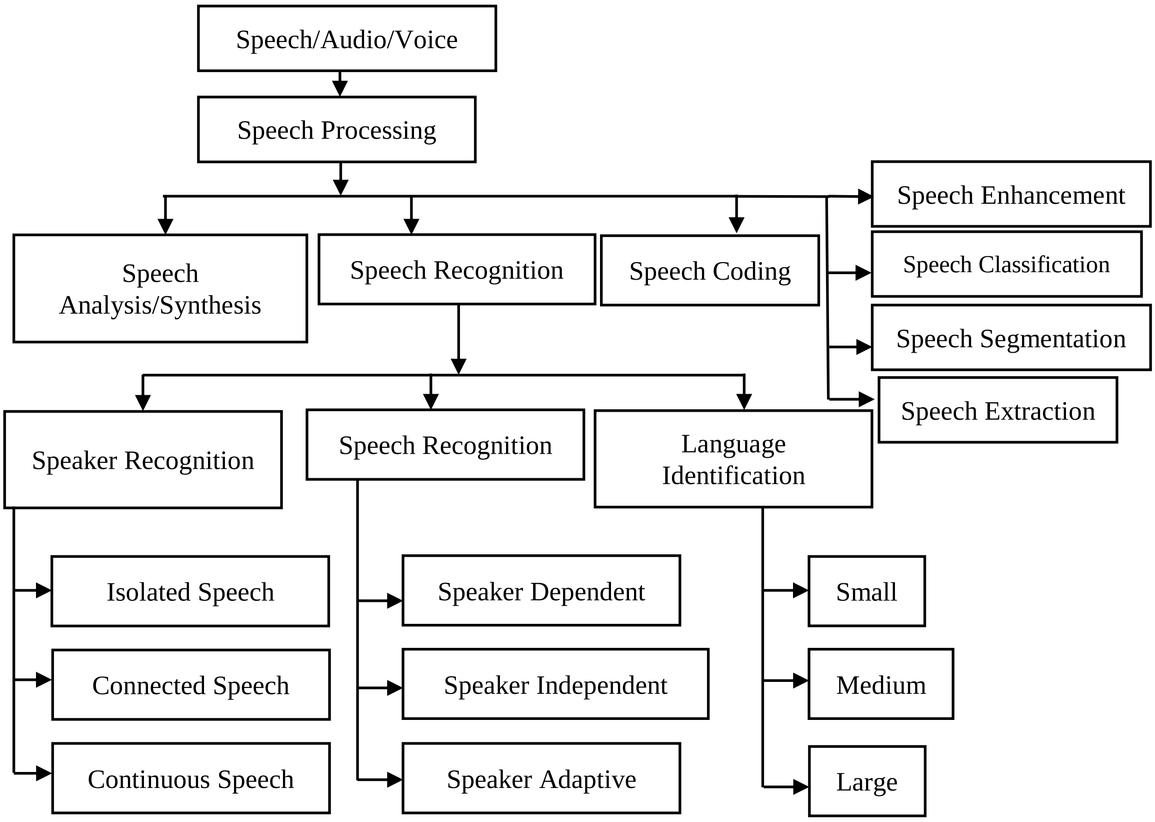

Fig 7: Classification of Speech based on Applications.

Classification of speech recognition is done based on different methods and algorithms used in their processing techniques and introducing more techniques and methods of enhancement to identify speech gradually improved the outcome. Speech Enhancement, Speech Classification, Speech Segmentation and Speech Extraction based on the features of Speech Recognizing techniques are the key identification goals listed in this paper [53-54], [78-79]. Recognition of speech is further distinguished as Recognition of Speakers and Language Identification which is achieved using a few techniques for speech processing. The recognition of the Speaker is further classified as Isolated Speech Connected Speech and Continuous speech. Isolated Speech is nothing but the exclusion of voice and voice information from the isolated area such as Speaker output devices or audio system components. Connected Speech is correlated with phonemes being formed into identifiable and meaningful voice data. Continuous Speech is used to classify the actual voice of the speaker and to translate the voice data into text data [65-67].

2. Literature Review:

There are various technologies and methods which are used to recognize and processing of speech which have their own methods and way of functioning to recognize the speech. The most popular approaches used in speech-related classifications in the following articles. In these, the trend-based and application-based publications are about vector machines for speech-based emotional recognition, age, gender, and speech-based speech recognition. They work based on Speech's emotion recognition.

The remaining publications are based on other Speech-based technology classifications [34-36]. The above table is distinguished based on the approach, and the method used by the proposed method to remove the maximum result errors. In this study, we will discuss different classifications of speech technologies and methods based on a few components that are segmentation, repositories, enhancement classification and extraction of features. These classifications will provide us with a better understanding of existing methods and technologies. These papers are based on Speech Enhancement and their methods and algorithms to improve speech more effectively and get the result based on the application and its uses. Now the most effective technique used to develop and enhance speech is the Pitch-adaptive short-term Fourier transformation (PASTFT), the use of ultrasonic doppler frequency changes induced by facial movements to improve high-acoustic noise-contaminated speech [56]. They are used to change voice-based applications. The most commonly used and reputed databases in this database category are TIMIT Corpus and Emo-DB Databases that are used in speech-based applications that have maximum accuracy in recognizing the words of different languages related to a specific use of Speech Applications. Further classifications are concerned with speaking systems and their applications.

Table.1: List of Speech Databases used for Speech processing

S.No.	Author	Databases	Results
1	Gabor Gosztolya [11], 2019	The Munich Bio voice, Hungarian Emotion, iHEARu-EAT Corpus and EGG Corpus for Cognitive Load	Both UARs (65.3%, 65.2%, and framework-level usability sets respectively) are significantly larger than most published results, which fell within the range of 63.1% –63.7%.
2	Anjali Bhavan.et.al [13], 2019	Emo-DB, RAVDESS, and IITKGP-SEHSC corpus	Comparing the proposed output of the system with those datasets was 92.45%, 75.69%, and 84.11%.
3	Johannes Stahl.et.al [14] ,2019	NOISEX-92 and TIMIT Core database	We also estimated a 0.80 real-time PACO, 0.22 PADDi, 0.23 COCA and 4.90 MVDR variables along with better than previously described algorithm settings.
4	NP Narendra.et.al [15], 2019	UA-speech database, Dysarthric, and TORGO	The accuracy of the definition of two sets of glottal parameters extracted from QCP and GIF ranges from 63 to 80%, coded for NB and WB, after the TORGO and UA Speech datasets.
5	Pravin Bhaskar Ramteke.et.al [16], 2019	IIIT Hyderabad Marathi, TIMIT and IIIT-H Indic Speech Databases	An accuracy of 95%, 96,87%, and 96,12% is obtained within the tolerance range of 10 ms.
6	Milton Sarria-Paja.et.al [17], 2018	CHAINS and TIMIT Database	Results showed of improvements of up to 79% and 60% for positive and whispered speech could be achieved, respectively, compared to a reference model equipped with I derived from cepstral coefficients of Mel frequency.
7	Yuan Luo.et.al [18],2018	Switchboard Database	The BNF WER (L2,1 overlapping group lasso) decreased by 5.59 percent compared to BNF (L1 lasso), and the WER decreased by 2.56 percent compared to BNF (L2 lasso sparse band).
8	Turgut Özseven.et.al [19], 2018	EMO-DB, EMOVA, Interface 05 and SAVEE Datasets	The performance achieved with SPAC is similar to the results obtained with a 1 percent -3 percent variation with other toolboxes. This is an example of the quality of SPAC feature sets.
9	Alexey Sholokhov.et.al [22], 2017	NIST SRE2010, RSR2015 and Red dots database	Compared to other contrastive SAD approaches, the proposed semi-supervised SAD is comparatively less contingent.
10	Athulya [21], 2018	TIMIT database	The combination of the classifier with the fused feature set significantly reduced error rates from 8.75 percent to 2.5 percent, which is much lower than that of regular PNCC classifiers.
11	Mathieu Labrunie et.al [23], 2018	First wide server for a French speaker with high-quality RT-MRI midsagittal pictures	The SPF method used 55 seconds to calculate the calculation times of an isolated image, and 32 seconds in succession, which is an RMS error of 0.93 mm.
12	Milton Sarria-Paja.et.al [20], 2017	TIMIT, CHAINS and wTIMIT Database	Relatively 66% and 63 percent improved on a reference system based on MFCCs and multi-conditional learning respectively for whispered and regular speech.
13	Erfan Loweimi et.al [8], 2017	AURORA-2 Database	The techniques are particularly useful in SNRs less than 10dB and return a complete increase in accuracy of more than 7% and 10% in SNRs between 0 and 5 dB.
14	Deepak Baby.et.al [9], 2015	AURORA version 2 and 4 databases.	DNN trained on noisy learning results in a comparative increase of around 40 percent over the GMM model and is even stronger than the most effectively retrained GMM set.

15	Pejman Mowlaee.et.al [24],2015	GRID corpus	In terms of mean and confidence interval, the median correlation means opinion score (CMOS) is shown on the right panel. The CMOS results show the choice that the combination of LSA and enhanced phase outperforms the noisy phase scenario between slightly better to better support the results expected by PSNR.

Table.2: List of Speech Enhancement based Evaluation methods

S.No.	Author	Methodology	
		Approach	Evaluation
1	Johannes Stahl. et.al [14], 2019	Model of signal and history, progression stage, phase of pitch adaptive analysis, factor K, and decomposition phase PASTFT	Study of interframe correlation, Sample autocorrelation series, frequency autocorrelation and frame-shift autocorrelation for Proper algorithm.
2	Ki-Seung Lee [26] , 2019	Ultrasound doppler detector Voice and USD correlation test, USD feature and Language-UDS comparison analysis feature	Speech prediction using UDS, DNN-based speech function estimated over-suppression of the noise for non-speech areas.
3	Emma Jokinen.et.al [27], 2017	Normal speech data conversion algorithm with Spectral tilt prediction from Speech data	Training Data Selection by Nonparallel Data & Process Analysis, Test Design, Participants based results and materials, methods for evaluation
4	Deepak Baby.et.al [9], 2015	Enhanced speech by adding dictionaries with Noisy Speech Model by NMF and Coupled Dictionary Process.	NMF Based Algorithm and DFT methods
5	Josef Kulmer.et.al [25], 2015	Issue description, ranking and meaning in-process and circulatory statistics by enhancing traditional speech and harmonic structure	Proposed phase estimation by phase decomposition, linear and temporary smoothing step removal and phase-enhanced speech synthesis
6	Pejman Mowlaee.et.al [24] ,2015	Model and background signal and problem description by phase and motivation traditional speech enhancement and harmonic structure	Synthesize the phase-enhanced speech signal through Binary Hypothesis Test Framework with SNR-based smoothing and phase smoothing using a hypothesis test

Table.3: List of Various Speech Segmentation Evaluation Methods

SNO	Author	Methodology	
		Approach	Evaluation
1	Pravin Bhaskar Ramteke.et.al [16], 2019	Datasets of TIMIT Corpus, IIIT-H Databases of speech-Marathi and IIIT-H Databases of speech-Hindi	Extraction based on spoken and unvoiced segmentation of the phoneme boundaries distinguishing in spoken and unvoiced regions
2	Laura J. Batterink.et.al [31], 2019	Participants of materials and methods stimulation and procedures	Behavioural Data Analysis and EEG Recording and Analysis
3	Mathieu Labrunie.et.al [23], 2018	Related research on the assessment of segmentation errors and the description of segmentation methods	Speech corpus and MRI in real-time, further corpus determination based on supervised machine learning methods
4	Okko Räsänen.et.al [32], 2017	Speech datasets and the role of syllables in speech perception of Sonority	An oscillator model for rhythmic segmentation based on sonority, Technical details based on tests based on various data sets
5	Toni Cunillera. et.al [33], 2016	Participants of resources and methods sensations and artificial language sources focused on the learning process of anchor words	Electrophysiological recordings and ERP data analysis
6	A. Stan.et.al [34], 2015	Related work on overview of the proposed method	Speech recognition with a skip network, acoustic model learning, trust evaluation.

Table.4: List of Various Feature Extraction methods of Speech

SNO	Author	Methodology	
		Approach	Evaluation
1	Gabor Gosztolya [11], 2019	EGG Corpus, iHEARu-EAT Corpus, Munich Bio Voice Corpus, Cognitive Speech Load and Hungarian Emotion Corpus.	Extraction by frame-level by Posterior-thresholding feature classification at the utterance level.
2	Xiao yuan Wang [35], 2019	The methodology focused on Moving emotion extraction method based on factor analysis, detailed assessment and emotional PAD model, identification model construction and data acquisition	Identification framework extraction of the function and calibration of the model
3	Michalis Papakostas.et.al [36], 2018	CNN's for Learning Dataset and Augmentation Audio Classification and Segment Representation Using CNNs to classify audio segments and details of implementation	Experimental appraisal focused on test data sets and results-based performance measures
4	Turgut Özseven.et.al [19] , 2018	Materials and methods of SPAC toolbox design	Pre-processing configurations, storage, post-processing, identification, and lot analysis functionality
5	Yuan Luo [18], 2018	BN Extraction Model	Overlapping group Lasso Sparse Deep Neural Network with Sparse group lasso model by BN Extraction Model and BN-DNN model

6	Takaaki Hori [37], 2017	The proposed system based on system overview, beamforming, LSTM speech improvement robust feature extraction, acoustic modeling, language modeling and system combination.	CHiME-3 task tests, baseline ASR performance, beamforming and voice enhancement with the effect of large-scale language models, acoustic model training with noisy multi-channel input, LSTM single-channel enhancement with noise-robust features and device combination

All these methods and techniques are based on machine learning and few of them are traditional and often used for speech recognition. They are

1. Hidden Markov Model (HMM)
2. Dynamic Text Warping (DTW)
3. Deep Neural Networks

Such techniques are used to learn and recognize speech, but also in future work, they have certain inconveniences and weaknesses which can help us to improve speaker technology quickly.

Below are the main issues of these speech recognition systems. HMMs Design Problems

- Several unstructured parameters still occur.
- First-order property is limited to their first-order property
- It cannot convey hidden state dependencies.
- The higher-order relationship between amino acids in a protein molecule cannot be identified.
- A fairly restricted HMM can only represent a small fraction of distributions over the space of potential sequences.

The technology of speech recognition and processing is evolving, increasing and improving. Technology for speech awareness can interact with other disabled persons. This allows the control of the digital system. Great opportunity in the future to extend the spoken network of engineering. Enhancing speech recognition can provide improved services to people with disabilities and provide our system with a secure environment with voice authentication. We are still well on the way before us because of the high level of competition on the market between this tech giant and the growing prevalence of companies jumping in to produce space content.

V. RESULT SECTION

The entire work evaluated on Python 3.8 with configuration of the system is Intel Core I5, 8GB RAMS, RADEON Card etc. Overall experiment performed on 3 benchmark dataset ie. Emo-DB, RAVDNESS and IITKGP-SEHSC Corpus. The proposed work used multi-layer Perceptron algorithm gives better result in term of accuracy 86.2%, 80.21 & 85.43 respectively Emo-DB, RAVDNESS and IITKGP-SEHSC Corpus dataset. Finally, it can be concluded that, overall accuracy improved as compared to existing technology as shown in Table 5 & Fig. 8-10.

Table.5: Comparison of Accuracy with Existing Technology

Dataset	Reference	Methodology	Accuracy (%)
Emo-DB	[3]	LBP	60
	[23]	Linear SVM	87.7
	[32]	SVM (Gaussian kernel)	88.9
	[51]	CNN	91.28
	[13]	Bagged SVM	92.45
	Proposed	**ANN**	**86.2**
RAVDESS	[4]	SVM	60.1
	[16]	DNN	64.52
	[38]	AdaBoost + SVM	72.10
	[13]	Bagged Ensemble + SVM	75.69
	Proposed	**ANN**	**80.21**
IITKGP-SEHSC Corpus	[6]	GMMs	73.68
	[38]	AdaBoost	77.19
	[13]	SVM	84.11
	Proposed	**ANN**	**85.43**

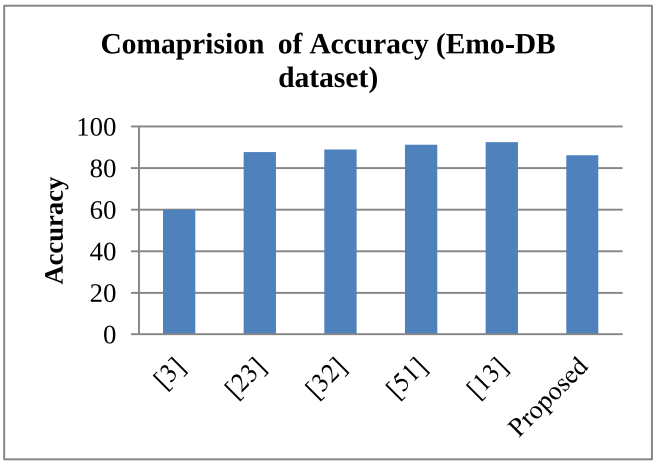

Fig. 8: Comparison of Accuracy with Existing Technology on Emo-DB dataset

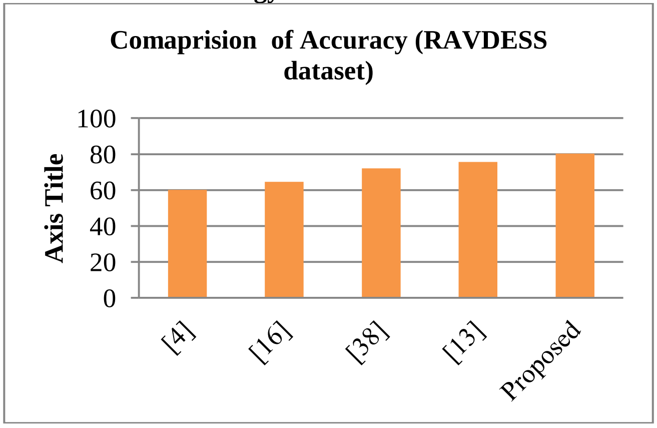

Fig. 9: Comparison of Accuracy with Existing Technology on RAVDASS dataset

Retrieval Number: E5715018520/2020©BEIESP
DOI:10.35940/ijrte.E5715.018520

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

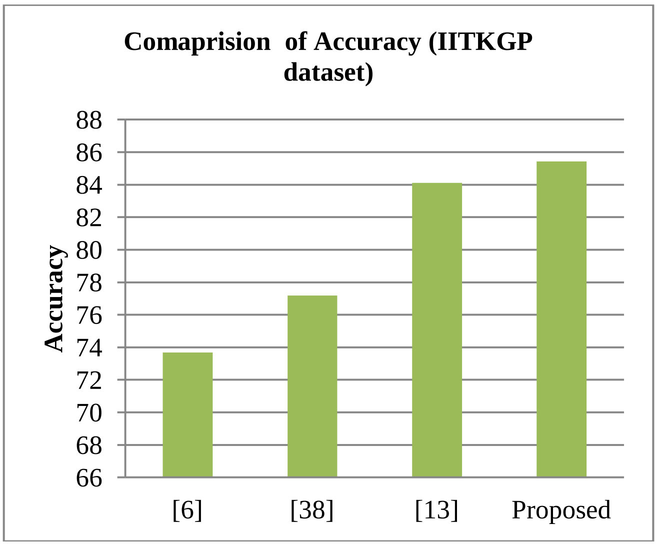

Fig. 10: Comparison of Accuracy with Existing Technology on IITKGP dataset

VI. CONCLUSION AND FUTURE WORK

We have listed and distinguished speaking technologies that are focused on specifications, databases, classification, feature extraction, enhancement, segmentation and process of Speech Emotion recognition in this paper.

We also integrated existing technologies and their use in specific technical areas, providing us with clear information on and developments of emerging speech-related technology on recognition of the emotion in speech. Some challenges need to be improved & one of them improved with the proposed work. The multi-layer Perceptron algorithm proves better result in term of accuracy 86.2%, 80.21 & 85.43 respectively Emo-DB, RAVDNESS and IITKGP-SEHSC Corpus dataset. Finally, it can be concluded that, overall accuracy improved as compared to existing module.

Future Scope: Systems linked to speech are rapidly changing and evolving. The early implementation of Speech Engineering has achieved different levels of success. The hope for the future is significantly higher quality in almost every area of speech-related technologies, with more robustness for speakers, ambient noise, etc. Such tools can be used as an application-based device and used in the medical field by recognizing the speech emotion and moreover, they can be used for home security purposes which can help the households and neighbours to provide emotional assistance to the people in need. This will ultimately create a secure and usable speech interfaces that are available for universal use for all telecommunications services.

REFERENCES

1. Errattahi, R., El Hannani, A., & Ouahmane, H, Automatic Speech Recognition Errors Detection and Correction: A Review. Procedia Computer Science, 128, 32–37(2018).
2. Ziheng Zhou, Guoying Zhao, Xiaopeng Hong, Matti Pietik¨ainen, A Review of Recent Advances in Visual Speech Decoding, Image and Vision Computing (2014).
3. Bhavan, P. Chauhan, Hitkul, et al., Bagged support vector machines for emotion recognition from speech, *Knowledge-Based Systems* (2019).
4. Saeid Safavi, Martin Russell, Peter Jancovi ˇ c, Automatic Speaker, Age-group and Gender Identification from Children's Speech, Computer Speech & Language (2018).
5. Partha Mukherjee, Soumen Santra, Subhajit Bhowmick et al., Development of GUI for Text-to-Speech Recognition using Natural Language Processing (2018).
6. Reza Sahraeian and Dirk Van Compernolle, *Member, IEEE*, Cross-lingual and Multilingual Speech Recognition Based on the Speech Manifold, IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 25, NO. 12, pp. 2301-2321(2017)
7. Rakhimov Mekhriddin Fazliddinovich, Berdanov Ulug'bek Abdumurodovich, Parallel Processing Capabilities in the Process of Speech Recognition, 978-1-5386-2168-4(2017).
8. Erfan Loweimi, Jon Barker et al., Statistical Normalisation of Phase-Based Feature representation for Robust Speech Recognition, 978-1-5090-4117-6, pp. 5310-5314 (2017).
9. Deepak Baby, Tuomas Virtanen et al., Coupled Dictionaries for Exemplar-Based Speech Enhancement and Automatic Speech Recognition. IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 23, NO. 11, pp. 1788-1799(2015).
10. Panikos Heracleous, Viet-Anh Tran, et al., Analysis and Recognition of NAM Speech Using HMM Distances and Visual Information, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 18, NO. 6 pp. 1528-1538(2010).
11. Songfang Huang and Steve Renals, Hierarchical Bayesian Language Models for Conversational Speech Recognition, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 18, NO. 8 pp. 1941-1954(2010).
12. Nicolae Duta, Richard Schwartz et al., Analysis of the Errors Produced by 2004, BBN Speech Recognition System in the DARPA EARS Evaluations, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 14, NO. 5, pp. 1745-1753(2006).
13. G. Gosztolya, Posterior-thresholding feature extraction for paralinguistic speech classification, *Knowledge-Based Systems* (2019).
14. Johannes Stahl, Pejman Mowlaee, Exploiting temporal correlation in pitch-adaptive speech enhancement, Speech Communication 111 (2019) 1–13 (2019).
15. N.P. Narendra, Paavo Alku, Dysarthric speech classification from coded telephone speech using glottal features, *Speech Communication* (2019).
16. Pravin Bhaskar Ramteke, Shashidhar G. Koolagudi, Phoneme boundary detection from the speech: A rule-based approach, Speech Communication 107 (2019) 1–17(2019).
17. Milton Sarria-Paja, Tiago H. Falk, Fusion of Bottleneck, Spectral and Modulation Spectral Features for Improved Speaker Verification of Neutral and Whispered Speech, *Speech Communication* (2018).
18. Yuan Luo, Yu Liu, Yi Zhang, Congcong Yue, Speech Bottleneck Feature Extraction Method Based on Overlapping Group Lasso Sparse Deep Neural Network, *Speech Communication* (2018),
19. Turgut Özseven, Muharrem Düğenci, SPeech ACoustic (SPAC): A novel tool for speech feature extraction and classification, Applied Acoustics 136 (2018) 1–8(2018).
20. Sandeep Kumar, Sukhwinder Singh and Jagdish Kumar "Automatic Live Facial Expression Detection Using Genetic Algorithm with Haar Wavelet Features and SVM" in Wireless Personal Communication Springer Journal (SCI) DOI: 10.1007/s11277-018-5923-y.
21. Sandeep Kumar, Sukhwinder Singh, and Jagdish Kumar, "Gender Classification Using Machine Learning with Multi-Feature Method" in IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, USA, January 7th-9th, 2019
22. M. Sarria-Paja, T. Falk, Fusion of auditory inspired amplitude modulation spectrum and cepstral features for whispered and normal speech speaker verification, Computer Speech & Language (2017).
23. Athulya, M.S., Sathidevi, P.S., Speaker verification from codec distorted speech for forensic investigation through a serial combination of classifiers, Digital Investigation (2018).
24. Alexey Sholokhov, Md Sahidullah, Tomi Kinnunen, Semi-Supervised Speech Activity Detection with an Application to Automatic Speaker Verification, *Computer Speech & Language* (2017).
25. Mathieu Labrunie, Pierre Badin et al., Automatic segmentation of speech articulators from real-time midsagittal MRI based on supervised learning, Speech Communication, pp. 27–46(2018).
26. Pejman Mowlaee and Josef Kulmer, Harmonic Phase Estimation in Single-Channel, Speech Enhancement Using Phase Decomposition and SNR Information, IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 23, NO. 9(2015).
27. Josef Kulmer and Pejman Mowlaee, Phase Estimation in Single-Channel Speech Enhancement Using Phase Decomposition, IEEE

SIGNAL PROCESSING LETTERS, VOL. 22, NO. 5(2015).

28. Ki-Seung Lee, Speech enhancement using ultrasonic doppler sonar, Speech Communication 110 (2019) 21–32(2019).
29. Emma Jokinen, Ulpu Remes et al., Intelligibility Enhancement of Telephone Speech Using Gaussian Process Regression for Normal-to-Lombard Spectral Tilt Conversion, IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 25, NO. 10(2017).
30. Li Wang, Xiaochu Liu, et al., Analysis and classification of hybrid BCI based on motor imagery and speech imagery, Measurement 147 (2019) 106842(2019).
31. S. Chowdhury et al., Automatic classification of speech overlaps Feature representation and algorithms, Computer Speech & Language (2018).
32. Marijn Huijbregts and Franciska de Jong, Robust speech/non-speech classification in heterogeneous multimedia content, Speech Communication 53 (2011) 143–153(2010).
33. Battering LJ, Paller KA, Statistical learning of speech regularities can occur outside the focus of attention, CORTEX (2019).
34. Okko Räsänen, Gabriel Doyle et al., Pre-linguistic segmentation of speech into syllable-like units, Cognition 171 (2018) 130–150(2017).
35. Toni Cunillera, Matti Laine et al., Headstart for speech segmentation: a neural signature for the anchor word effect, Neuropsychologia 82 (2016) 189–199(2016).
36. Stan, A., et al., ALISA: An automatic lightly supervised speech segmentation and alignment tool. Comput. Speech Lang. (2015).
37. Xiaoyuan Wang, Yaqi Liu et al., Feature extraction and dynamic identification of driver's emotions, transportation Research Part F 62 (2019) 175–191(2019).
38. Michalis Papakostas, Theodoros Giannakopoulos, Speech-Music Discrimination Using Deep Visual Feature Extractors, *Expert Systems with Applications* (2018).
39. Chao Sui, Roberto Togneri, Mohammed Bennamoun, A Cascade GrayStereo Visual Feature Extraction Method For Visual and Audio-Visual Speech Recognition, *Speech Communication* (2017).
40. Charles Jankowski, Ashok Kalyanwamy et al., NTIMIT: A Phonetically Balanced, Continuous Speech, Telephone Bandwidth Speech Database, C€I2847-2/90/0000-010, pp. 109-112(1990).
41. John H.L. Hansen and Sahar E. Bou-Ghazale, Getting Started with SUSAS: A Speech Under Simulated and Actual Stress Database, EUROSPEECH '97 5th European Conference on Speech Communication and Technology Rhodes, Greece, September 22-25, (1997).
42. C. Breazeal, L. Aryananda, Recognition of affective communicative intent in robot-directed speech, Autonomous Robots 2 (2002) 83-104.
43. T. Nwe, 8. Foo, L. Be Silva, Speech emotion. recognition using hidden Markov models, Speech Commun. 41 (2003) 603-623.
44. M. Slaney, G. McRoberts, Baby ears: a recognition system 'for affective vocalizations, Speech Commun. 39 (2003) 367--3 84.
45. B. Schgller, S. Reiter; R. M01191, M..Al-Hames, M. Lang, G. R}goll, Speaker independent speech emotion recognition by ensemble classification, in IEEE International Conference on Multimedia and Expo, 2005. ICME 2005, 2005, pp. 864-867.
46. J. Zhou, G. Wang, Y. Yang, P. Chen, Speech emotion recognition based on rough set and SVM, in 5th IEEE International Conference on Cognitive Informatics, 2006, 1CC12006, vol. 1, 2006, pp. 53-61.
47. J. Kominek and A. Black, "The CMU ARCTIC databases for speech synthesis," Tech. Rep. CMU-LTI-03-177, Language Technologies Institute, Carnegie Mellon University, 2003.
48. Tanja Schultz, Michael Wand et al., Biosignal-Based Spoken Communication: A Survey, IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 25, NO. 12, 2017, pp. 2257-2271.
49. Lucy Liao, Mark A Gregory, Algorithms for Speech Classification, WIT University Melbourne Victoria Australia 3001, pp. 623-627,1999.
50. Alaa Ehab Sakran, Sherif Mahdy Abdou, A Review: Automatic Speech Segmentation, International Journal of Computer Science and Mobile Computing, Vol.6 Issue.4, April- 2017, pg. 308-315.
51. Shreya Narang, Ms. Divya Gupta, Speech Feature Extraction Techniques: A Review, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.3, March- 2015, pg. 107-114.
52. Sandeep Kumar, Rohit Raja, and Archana Gandham, "Tracking an Object using Traditional MS (Mean Shift) and CBWH MS (Mean Shift) Algorithm with Kalman filter" in the Springer Nature Book Series "Algorithm For Intelligent Systems-Element of Statistical Learning" (Accepted).
53. Sandeep Kumar, Sukhwinder Singh, and Jagdish Kumar, "A Multiple Face Detection Using Hybrid features with SVM Classifier" in the Springer Nature on Data Communication and Networks with ISBN: 978-981-13-2254-9
54. Sandeep Kumar, Sukhwinder Singh and Jagdish Kumar "Live Detection Of Face Using Machine Learning with Multi-Feature Method" in Wireless Personal Communication Springer Journal (SCI) DOI: 10.1007/s11277-018-5913-0.
55. E. A. Kaur and E. T. Singh, "Segmentation of continuous Punjabi speech signal into syllables," in Proceedings of the World Congress on Engineering and Computer Science, vol. 1. Citeseer, 2010, pp. 20–22.
56. S. Ratsameewichai, N. Theera-Umpon, J. Vilasdechanon, S. Ua-trongjit, and K. Likit-Anurucks, "Thai phoneme segmentation using dual-band energy contour," ITC-CSCC: 2002 Proceedings, pp. 111–113, 2002.
57. B. Zio´ łko, S. Manandhar, R. C. Wilson, and M. Zio´ łko, "Wavelet method of speech segmentation," in Signal Processing Conference, 2006 14th European. IEEE, 2006, pp. 1–5.
58. M. Tolba, T. Nazmy, A. Abdelhamid, and M. Gadallah, "A novel method for Arabic consonant/vowel segmentation using the wavelet transform," International Journal on Intelligent Cooperative Information Systems, IJICIS, vol. 5, no. 1, pp. 353–364, 2005.
59. J. Kamarauskas, "Automatic segmetation of phonemes using ar- tificial neural networks," Elektronika ir Elektrotechnika, vol. 72, no. 8, pp. 39–42, 2015.
60. Y. Suh and Y. Lee, "Phoneme segmentation of continuous speech using multi-layer perceptron," in Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on, vol. 3. IEEE, 1996, pp. 1297–1300.
61. M. M. Rahman, F. Khatun, and M. A.-A. Bhuiyan, "Blocking black area method for speech segmentation," Editorial Preface, vol. 4, no. 2, 2015.
62. A. Hossain, N. Nahid, N. N. Khan, D. C. Gomes, and S. M. Mugabe, "Automatic silence/unvoiced/voiced classification of Bangla velar phonemes: New approach," 8th ICCIT, Dhaka, 2005.
63. M. Kalamani, S. Valarmathy, and S. Anitha, "Hybrid speech segmentation algorithm for continuous speech recognition."
64. S. Nishi and S. Parminder, "Automatic segmentation of wave file," Int J of Comput Sci Commun, vol. 1, no. 2, pp. 267–270, 2010.
65. P. Bansal, A. Pradhan, A. Goyal, A. Sharma, and M. Arora, "Speech synthesis-automatic segmentation," International Journal of Computer Applications, vol. 98, no. 4, 2014.
66. J. D. Markel, **A.** H. Gray, Jr., Linear Prediction of Speech. New York: Springer Verlag, 1976.
67. L. Liao "Multi-Category CELP Coder with Dynamic Vector Quantisation". RMIT, Master Thesis, April 1999.
68. L. R. Rabiner, M. R. Sambur, "An Algorithm for Determining the Endpoints of Isolated Utterances", The Bell System Technical Journal, Vol. 54, No. 2, pp. 297-315, Feb.1995.
69. R. *0.* Duda, P. E. Hart, Pattern Classification and Scene Analysis. Jone Wiley & Sons, 1973.
70. J. Benesty, S. Makino, J. Chen (ed). Speech Enhancement. pp.1-8. Springer, 2005. ISBN 978-3-540-24039-6.
71. J. Benesty, M. M. Sondhi, Y. Huang (ed). Springer Handbook of Speech Processing. pp.843-869. Springer, 2007. ISBN 978-3-540-49125-5.
72. J. Benesty, M. M. Sondhi, Y. Huang (ed). Springer Handbook of Speech Processing. Springer, 2007. ISBN 978-3-540-49125-5.
73. J. Benesty, S. Makino, J. Chen (ed). Speech Enhancement. Springer, 2005. ISBN 978-3-540-24039-6.
74. P. C. Loizou. Speech Enhancement: Theory and Practice. CRC Press, 2013. ISBN 978-1-466-50421-9.
75. Celso Auguiar, in CCRMA - Center for Computer Research in Music and Acoustics. Stanford University on Modelling the Excitation Function to Improve Quality in LPC's Resynthesis**.**
76. Sahidullah, Md.; Saha, Goutam (May 2012). "Design, analysis and experimental evaluation of block-based transformation in MFCC computation for speaker recognition". Speech Communication 54 (4): 543–565.
77. Hynek Hermansky, Eric A. Wan, and Carlos Avendano, Oregon Graduate Institute of Science & Technology Department of Electrical Engineering and Applied Physics, Speech enhancement based on temporal processing.
78. H. Hermansky and N. Morgan, Rasta processing of speech, IEEE Trans. on Speech and Audio Processing, vol. 2, no. 4, pp. 578-589, 1994.
79. Liang Lu, Member, IEEE, and Steve Renals, Fellow, IEEE, IEEE SIGNAL PROCESSING LETTERS, Probabilistic Linear Discriminant Analysis for

Acoustic Modelling, VOL. X, NO. X, 201.

80. D.O. Shaughnessy, Speech Communication: Human and Machine. Second Edition India: University Press (India) Private Limited, 2001
81. H.Hermansky, B. A. Hanson, H. Wakita, "Perceptually based Linear Predictive Analysis of Speech," Proc. IEEE Int. Conf. on Acoustics, speech, and Signal Processing, pp. 509-512, Aug.1985
82. H. Hermansky, B. A. Hanson, and H. Wakita, "Perceptually based Processing in Automatic Speech Recognition," Proc. IEEE Int. Conf. on Acoustics, speech, and Signal processing, pp. 1971-1974, Apr.1986.
83. J. S. Garofalo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, ―DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM‖, Tech. rep., U.S. Dept. of Commerce, NIST, Gaithersburg, MD, (1993).
84. V. Zue, S. Seneff and J. Glass, ―Speech database development at MIT: TIMIT and beyond‖, Speech Communication, vol. 9, no. 4, (1990), pp. 351-356.
85. Fisher, William M.; Doddington, George R., and Goudie Marshall, Kathleen M. (1986). The DARPA Speech Recognition Research Database: Specifications and Status. pp. 93-99.
86. Meigrgwiyer, 8., Merlin, T., Blouet, R., and Bonastre, J.-F. (2002‘). NIST 2002 speaker recognition evaluation: LIA results. In Proceedings MST 2002 Speaker Recognition Workshop, Vicnna, Virginia.

AUTHORS PROFILE

Mr. C. Andy Jason is presently pursuing M.tech in the Department of Electronics & Communication Engineering, Sreyas Institute of Engineering & Technology, Hyderabad, India. He did his B.tech in the Department of Electronics & Communication Engineering, Anurag Group of Institutions, Hyderabad, India. His area of research includes Speech Recognition and Machine Learning. He also attended number of seminars, workshops and conferences.

Dr. Sandeep Kumar is presently working as a Professor in the Department of Electronics & Communication Engineering, Sreyas Institute of Engineering & Technology, Hyderabad, India. He has good Academics & Research experience in various areas of Electronics and Communication. His area of research includes Embedded systems, Image processing, and Biometrics. He has filed successfully 7 National & 1 International Patent. He has been received 3 times invitation being a Guest in Scopus Indexed IEEE/Springer Conferences. He has been invited 4 times being an expert in various Colleges/universities. He has published 70 research papers in various International/National Journals (including IEEE, Springer etc.) and Proceedings of the reputed International/ National Conferences (including Springer and IEEE). He has been awarded “Best Paper Presentation” in Nepal & India respectively 2017 & 2018. He has been awarded for “Best Performer Award” in Hyderabad, India, 2018. He has been awarded also “Young Researcher Award” in Thailand, 2018. He has been awarded also “Best Excellence Award” in Delhi, 2019. He is an active member of 17 various Professional International Societies. He has been nominated in the board of editors/reviewers of 23 peer-reviewed and refereed Journals (including IEEE, Springer). He is also attended 24 seminars, workshops and short-term courses in IITs etc.