Call Transcription and Text Classification - A Review

Aiswarya Lakshmi Thakka Ravunniy UG student Dept. of ECE NSSCE Palakkad	**Anaswara G. K.** UG student Dept. of ECE NSSCE Palakkad	**Apoorva Eliza John** UG student Dept. of ECE NSSCE Palakkad
Dhanusha R. UG student Dept. of ECE NSSCE Palakkad	**Gayathry Mohan** UG student Dept. of ECE NSSCE Palakkad	**Sreelekshmi P. S.** Assistant professor Dept. of ECE NSSCE Palakkad

ABSTRACT
A common method of data exchange is speech. Keeping track of every word exchanged through verbal means is difficult. Mostly, verbal communication is carried out through voice calls. For keeping tab of information exchanged, calls can be converted to text. These text files can be classified according to the content of the document. Also, unwanted call recordings have to be eliminated. This research contains the various methods required to implement call transcription and text classification. More than 50 papers were reviewed regarding the sections, call recording, transcription and ordering of text documents.

Keywords
Call recording; speech to text; classification; feature extraction; HMM; Baye's classifier

1. INTRODUCTION
In today's world, vast amount of information is exchanged; the most common medium of exchange being voice calls. Keeping tab of every word or phrase in a conversation is impossible, which results in loss of information. These lost chunks may contain significant data, the loss of which may lead to many repercussions. This necessitates the need for keeping track of information passed through voice calls. Call recording is usually implemented in call centers [1][2] and by service providers [3], also the call can be recorded by the user themselves. These recorded calls can be later reviewed [4]. But in this fast-paced world, rather than wasting time reviewing each call, if these calls are categorized and ordered according to the value of the content in them it will be very beneficial to the user. Also, rather than hearing this audio again, if there is a transcript, i. e, a textual form of the call content it would be highly time-saving. The conversion of audio or video forms into text format for later reference is called transcription. After conversion, a vast amount of storage cannot be wasted to store every call. It has to be determined whether a particular call has some important information, or whether the content in the call consists of any valuable information to the user at all.[1]

Call transcription is the process of converting voice or video calls in to written words. These are stored as plain text in conversational language. It can either be done in real time or using past conversation records. Call transcription is a powerful tool for business, medical and training sectors since the text file has better analysis features available than that for audio files.

In earlier days transcription was a tedious job as speech had to be written down using advanced skills like shorthand. Also it has to be done in real time. With the invention of tape cassettes and recorders it became much easier to record the data and send it to a different location for transcription

Now, with the advancements in technology, transcription can be done very rapidly in an effective manner. Recorded calls can now be uploaded to a server[5][6][7] wherein it will be automatically converted to the corresponding text. It is done by extracting the features of the audio file. An issue concerning it was the lower accuracy for nonnative speakers of a language. Main reason is because pronunciations of the non-native speakers differ from those of native speakers. Thus acoustic model adaptations were developed for the recognition of non- native speech. Two commonly used model for acoustic model adaptation are maximum likelihood linear regression (MLLR) and maximum a posterior (MAP) estimation.[8][9]

Acoustic model adaptations were limited to the cases where the identity of the speakers were known or the duration of the speech were only for a few minutes[10]. The language model adaptations then became significant that captures idiolect, which is the linguistic term referring to the variety of language used by a particular individual. Speaker idiolect provides discriminative information for speaker identification. LM adaptation is performed using common approach of linear interpolation. [9][11]

Speech recognition techniques were earlier used for assisting effective bidirectional communication of intensions between people of normal hearing and hearing-impaired [12].This requires speech to text conversion in real time.Real time speech to text conversion can be defined as conversion of words instantly after speaking [13]. It can also be used to support multilingual communication between people[14].Another area where it was used is for mammo class.

Mammo class is a web tool that allows users to enter a small set of variable values that describe a finding in a mammography, and produces a probability of this finding

being malignant or benign. The tool requires that the user types in every variable a value in order to perform a prediction. A speech-to-text interface integrated to Mammo Class that allows radiologists to speak up a mammography report instead of typing it in. This new Mammo Class module can take audio content, transcribe it into written words, and automatically extract the variable values by applying a parser to the recognized text. Results of spoken mammography reports show that the same variables are extracted for both types of input: typed in or dictated text [15].

Text classification uses several tools from information retrieval and machine learning[16][17]. Classification techniques have been applied in various fields such as spam filtering, e-mail routing, language identification, sentimental analysis etc. [18]

More than 50 papers were taken related to call recording, speech to text conversion and text classification and they were reviewed. The paper is organized as follows. Section 2 gives the collected data on call recording applications in an Android-based operating system. Section 3 gives the review of the architecture of speech to text conversion and various methods that can be used for each stage. Section 4 gives the review of different methods of text classification and categorization and Section 5 gives a review of recommendation algorithms, Section 6 gives an analysis on the most favourable methods in each section, Section 7 concludes the paper and suggests possible future work.

2. CALL RECORDING

The first step in call transcription is call recording. Call recording is the process of recording the calls using different applications.

Call recording software is used to record telephone conversations over PSTN or VoIP in the digital audio file format. Call recording applications developed in androids may not only record conversations but there are preferences for a user to record the calls in their choice of format. [19].These kinds of applications can be developed by using the latest technology such as SIP. They are scalable and can be implemented by telecom service provider companies, which works on VoIP [3]. An android application which is installed on the user's smartphone is used to record the calls. Then these calls are converted to the preset audio format and send to a designated server.[19][20][21]

Using call recording organizations have a means of auditing employee activities[22]. Call recordings can be used to manage risks by identifying the root cause of performance issues or keep track of performance[23]. It can safeguard business against unwarranted claims of wrongdoing helping to mitigate or altogether avoid costly legal actions. It can also be used to appeal to the customers by providing them products exactly according to their demands.

One of the main issues faced during call recording is the background noises. This sometimes meddles with speaker's voice. The removal of noise is formulated as a digital filtering system. Clean speech estimation can be obtained by passing the noisy speech through a linear filter. This sometimes leads to complete failure of voice communication. Some noise suppression techniques are noise proof microphone technology, headphone technology, compandor technology etc. Non-linear acoustic echo cancellation technology can be used for achieving anechoic communication. Compandor technology reduces the external noise while transmitting and reduce the background noise during the reception.[24][25][26]

Acoustic echo occurs due to large transmission delays. Echo cancellation and noise reduction can be achieved by conventional adaptive echo canceller along with a second adaptive filter in the sending path [27][28][29][30].

3. SPEECH TO TEXT CONVERSION

Speech is the basic mode of communication and the most natural and efficient form of exchanging information in human society. Speech to text conversion system is used to convert an audio file or a speech input from the microphone to text. Speech recognition system can be categorized into different classes based on the type of utterance that can be recognized [31]. Automatic speech recognition has two phases. The first one is training phase. In this phase, the system learns the reference pattern representing the different speech sounds. The second one is the recognition phase in which an unknown input pattern is identified by the set of references.

Speech to text conversion process can be broadly divided into three parts. They are endpoint detection, feature extraction and conversion to text. Endpoint detection method is used to detect the starting and end points of a word in a noisy environment. This method is based on the short-term log energy and short-term zero crossing rate [32]. Feature extraction is used to identify the useful components of speech signals that are used to identify the linguistic content and discard all other stuff which carries information like background noise, emotions etc. Various methods are used for feature extraction [33]. Most commonly used methods are Linear Predictive Coding (LPC)[34], Mel Frequency Cepstral Coefficient (MFCC) analysis, Wavelet Packet Decomposition (WPD)[34] and Perceptual Linear Prediction (PLP)[35]. MFCC method for feature extraction mimics the human auditory system and gives better performance rate. Also, MFCC being considered as frequency domain features are more accurate than the time domain features.

Steps involved in MFCC extraction are pre-emphasis, framing, windowing, Discrete Fourier Transform Computation, Mel frequency filtering, logarithmic transformation and Discrete Cosine Transform computation. Pre-emphasis is applied to spectrally flatten the input speech signal. Framing is blocking of speech samples obtained after analog to digital conversion of the spoken word into a number of frame signals with a frame length of 20 to 40 ms. Loss of information can be avoided by using overlapping. Windowing function is used to reduce the discontinuities at the start and end of the frame or to be smooth of the first and last points in the frame. DFT is used to convert the signals into frequency domain for more precise calculations. Mel frequency filtering uses a perceptual scale that helps to simulate the way human ear works. It corresponds to less resolution at high frequency and better resolution at low frequency. To make feature extraction speaker independent, Mel scale conversion is followed by logarithmic transformation. DCT is used to convert the Mel filtered spectrum back into the time domain since the Mel frequency coefficients are used as the time index in recognition stage [36] [37][38] [39]

After extracting the features the next step is to convert the speech into text. Various techniques are used for this purpose. The most commonly used methods are Hidden Markov Model (HMM), Dynamic Time Warping (DTW) [39], Vector Quantization (VQ) [40] and Artificial Neural Networks (ANN) [41][42][43]. Among the four methods HMM is the most acceptable and popularly used technique for speech to text conversion. The reasons for the popularity of HMM are

that they can be trained automatically and are simple and computationally feasible. The recognition rate and accuracy are very high for HMM. In HMM stochastic model is created considering the known utterance and then compares the probability that the unknown utterance was generated by each and every model [44] [45].Minimum distance classifier (MDC) and support vector machine (SVM) method can also be used for speech to text conversion [36]. Kalman filter can be used in speech recognition to remove the unwanted noise and to get filtered output. Kalman filtering is a state estimator that produces an optimal estimate and minimizes the mean square error [46]. The utterance-level Permutation Invariant Training (uPIT) is a practical technique for speaker independent multi-talker speech separation [47].

Sphinx is a continuous-speech, speaker-independent recognition system making use of hidden Markov acoustic models (HMMs). Sphinx features feasibility of continuousspeech, speaker-independent large-vocabulary recognition. CMU Sphinx uses MFCC features with noise tracking and spectral subtraction for noise reduction [48]

4. TEXT CLASSIFICATION
Text categorization (or text classification) is the allotment of a category to text documents according to the content of the document. Text classification has been studied in various areas like data mining and is also used in various domains like document organization and so on. It aims to assign predefined classes to the text document.[37][49][50] Text classification can be done either using the traditional approaches using feature extraction [37][49][51][52], by deep learning approaches[53] or by inductive learning methods [54].Text classification is done through the following steps: document representation or preprocessing, vector representation, categorization [55][56].

Some of the traditional methods include:

4.1 Decision Tree Algorithm
In Decision tree algorithm documents are categorized by moving through a tree structure where the nodes represent a question and leaves represent a category.[50]

4.2 K-Nearest Neighbor
K-nearest neighbor (k-NN) [37][49][52] is used to test the degree of similarity between documents and k training data and to store a certain amount of classification data, thereby determining the category of test documents[54]. An improved version of this is adaptive k-NN [57].

4.3 Naïve Baye's Classifier
Naive Baye's classifier is the simplest and the most widely used classifier. It models the distribution of documents in each class using a probabilistic model assuming that the distributions of different terms are independent of each other. There are two main models commonly used for Naive Bayes classifications. Both models aim at finding the posterior probability of a class, based on the distribution of the words in the document. The main difference between these two models is, one model takes into account the frequency of the words whereas the other one does not. These are Multi-variate Bernoulli Model and Multinomial Model [49][37]

4.4 Vector Based Methods
Some of the Vector-based Methods are Centroid Algorithm[37] and SVM [52][58][59]. In Centroid Algorithm, during the learning stage, only the average feature vector for each category is calculated and set as centroid-vector for the

category. A new document is easily categorized by finding the centroid-vector closest to its feature vector, while SVM (Support Vector Machine) is a form of Linear Classifiers that try to find a "good" linear separator between various classes. One advantage of the SVM method is that, it is quite robust to high dimensionality, i.e. learning is almost independent of the dimensionality of the feature space. It rarely needs feature selection since it selects data points (support vectors) required for the classification.

4.5 Deep Learning Approaches
Various Deep learning approaches include Artificial Neural Networks that are constructed from a large number of elements with an input an order of magnitudes larger than in computational elements of traditional architectures.[60] or Convolution Neural Networks which are very similar to ordinary neural networks as both of them are made up of neurons that have learnable weights and biases. CNNs have three main types of layers, namely, convolutional layer, pooling layer, and fully connected layer. [61] Also Fuzzy Correlation(fuzzy k-NN) can be used to deal with fuzzy information or incomplete data, and also convert the property value into fuzzy sets for multiple document classification. They can be used for multiclass text classification using one against one fuzzy support vector machines. Finally, Genetic algorithm aims to find optimum characteristic parameters using the mechanisms of genetic evolution and survival of the fittest in natural selection. It makes it possible to remove misleading judgments in the algorithms and improve the accuracy of document classification.[62]

4.6 Clustering
Similarly, clustering can also be used for grouping of text documents. It is the task of finding groups of similar documents in a collection of documents. The similarity is computed by using a similarity function.

Text clustering can be in different levels of peculiarity where clusters can be documents, paragraphs, sentences or terms. Different clustering algorithms can be used for grouping text documents. The text document can be represented as a binary vector, i.e. considering the presence or absence of a word in the document. Text clustering algorithms are split into many different types such as Hierarchical clustering algorithms that build a group of clusters which can be depicted as a hierarchy of clusters. [52][63] or k-means clustering which partitions in documents in the context of text data into k clusters. The main disadvantage of k-means clustering is that it is indeed very sensitive to the initial choice of the number of k. [49]

5. RECOMMENDATION ALGORTHM
Recommendation Algorithms can be used to classify the text documents in priority order. Most recommendation algorithms aggregate similar items and eliminates items that are already present. These algorithms are currently used for recommending some service[64] (social networking -friend lists, movies, shopping items etc) based on previous data or history of use or some priority set by the programmer. In [65] the authors talk about Friend Recommendation based on the proportion of Common Friends, based on user based collaborative filtering, normal check-in locations and have finally Friend recommendation based on Location. In [66] an improved item based recommendation algorithm is implemented for movie recommendation based on the genres watched previously. In [67] item to item based collaborative recommendation is used for recommending items in shopping lists. In [68] Internet of Things Service Recommendation System is used for drawing user's attention to new services. In

[69] the authors implement Multi Objective Evolutionary Algorithm for Pattern Mining (MOPM) for finding a set of optimal solutions.

6. ANALYSIS

In the previous sections, the various methods that can be used for implementing call transcription and text classification was explored.. The first part to be implemented is call recording. As reviewed, call recording can be done on the users Smartphone by developing an android application using the latest technology such as SIP and VoIP. Developing an application on the Android platform is the most affordable for the recording of audio in the desired format.

For the speech to text conversion part, first of all, the different feature extraction methods were reviewed.

It was found that LPC, PLP and MFCC were the most commonly used features extraction techniques in the field of speech recognition. The PLP technique is more adapted to human hearing, in comparison to the Linear Prediction Coding (LPC), because LPC employs linear computational technique whereas PLP uses transformed spectral characteristics of the audio to match human auditory system. The main difference between PLP and LPC is that the LP model assumes the-pole transfer function with a specified number of resonances within the analysis band. Also, when compared to PLP, LP approximates power distribution equally at all frequencies of the analysis band. The most efficient and prominent method used to extract spectral features is calculating Mel-Frequency Cepstral Coefficients (MFCC). It uses a nonlinear frequency scale which approximates the behavior of the auditory system. These coefficients are robust and adaptable to variations according to speakers and recording conditions. Of these, MFCC feature extraction is observed to be the most favourable, mainly because, MFCC feature extraction is working in the frequency domain and also because it can imitate the human auditory system.

While considering the various speech to text conversion methods, it has been observed that in noisy environment, the recognition performances of HMM is better than DTW ,also HMM is the most sophisticated and efficient algorithms for training and decoding. These factors have made HMM to be widely used in speech conversion applications. Another technique is vector quantization, but due to the inability to handle large vocabulary and high computational cost in clustering its use is limited. The HMM technique has significantly reduced computational cost and has been used for large vocabulary. The ANN is convenient for problems with limited information regarding characteristics of the input. However, the optimal ANN architecture to solve a problem must be selected by trial and error, which is a drawback. So, due to its predictive nature, HMM is observed to be the most favorable. As reviewed, CMU Sphinx tool, employs speech to text conversion using MFCC feature extraction and HMM modelling. Also, this tool implements an improved MFCC which removes noise using spectral subtraction. Hence, this tool has been observed to be the most favorable one for implementing the speech to text conversion part.

Inspecting the various text classification methods that have been reviewed, the following can be noted. One of the techniques used is neural networks which are difficult to interpret and accurate classification using networks are largely dependent on the architecture of network. Though it works well in noisy environment, trial and error method is used for deciding the number of layers and units in these layers.KNN is a text classification method which cannot be used until the

sample of each neighbor are available. Hence it's a time consuming process when compared to other methods. Since Naïve Bayes algorithm works on the principle of Bayes theorem, they give probabilistic predictions. It can handle continuous and discrete data and are not sensitive to irrelevant features. Since it is relatively robust, fast, and accurate it is more widely used.

Ideally it would be desired to order the text documents in descending order of priority, where priority is defined by the quality of content or information in each document. That is to say, methods that use only keywords to classify documents are not favorable as the same keyword can be used in any context. For example, a sentence like "I like to watch sports channels" has very low priority in terms of content compared to a sentence like "The most watched sport by Indians can be said as Cricket". If the keyword was chosen as 'sport' then both these sentences would have the same priority, but in terms of quality of information, both are not same. Hence this leads to the conclusion that a context-driven classification method would be the most favorable. Of the various methods that was reviewed, the Naive Bayes' Classifier looks to be the most favorable method as it looks at context by looking at a group of words to set a priority.

To justify the conclusion made above a comparison test was done among all these algorithms and results were obtained. [fig.1] shows the results obtained in the comparison test. Score shows the strength of prediction for different text classification algorithms. From this it is clear that Naïve Bayes classifier is better when compared to other techniques.
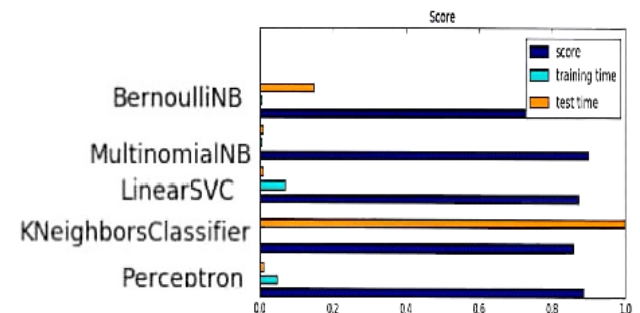

Fig.1 comparison of classification algorithms

Also literature for recommendation algorithms to learn about setting priority to text documents based on user's daily routine was reviewed. Hence, the order of priority can be changed by the program itself by reviewing the activities of the user on a daily basis.

7. CONCLUSION

This paper contains a review of the different existing technologies for implementing call transcription and text classification. First, this research reviews technique such as SIP or VoIP to record calls on android platform. Also, different speeches to text conversion methods were reviewed and MFCC feature extraction and HMM modelling was found more appropriate. Similarly, various text classification algorithms and recommendation algorithms were reviewed. In text classification algorithms Naïve Baye's algorithm had favourable results.

In future, call transcription and text classification can be adopted for creating a system where a user's call data can be saved as text documents in an order of priority. If implemented, such a system could prove to be useful for

keeping a reference of the data exchanged in user's everyday life. Also, user can have a log of all the important information [70][71] that was ever exchanged. This paper reviews methods to implement this idea that can be incorporated into the present systems for keeping a tab of every word exchanged through voice calls.

8. REFERENCES

[1] Christophe Vaudable, Laurence Devillers, ``Negative Emotions Detection As An Indicator Of Dialogs Quality In Call Centers'', *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2012

[2] Boaz M, Ben David and Michal icht, ``Voice changes in real speaking situations during a day, with and without vocal loading : assessing call center operators", *Journal of Voice,* Vol 30, April 2015

[3] Shrikant Subhash Warghade, Karthikeyan B, Suraj Rodrigues, Nagaraj Hebbar, ``Call recording solution using SIP", *Advances in Computational Sciences and Technology* Volume 10, Number 8 (2017)

[4] Fahim Hasan Khan, Mohammed Eunus Ali, Himel Dev, "A Hierarchical Approach for Identifying User Activity Patterns from Mobile Phone Call Detail Records'', *IEEE International Conference Netwroking Systems and Security*, 1-6, 2015

[5] Payman Mohassel, Ostap Orobets, and Ben Riva, ``Efficient ServerAided 2PC for Mobile Phones" *Proceedings on Privacy Enhancing Technologies* , 2016

[6] Lien-Wu Chen, Yu-Fan Ho, Wei-Ting Kuo and Ming-Fong Tsai, ``Intelligent file transfer for smart handheld devices based on mobile cloud computing", *international journal of communication systems,* 2015

[7] Suman Bhattacharjee, Sourav Kanta, Saket Modi, Madhumita Paul and Sipra DasBit, ``Disaster Messenger: An Android based Infrastructure Less Application for Post Disaster Information Exchange", *IEEE International Conference on Advanced Networks and Telecommunication Systems,* 2016

[8] Zhirong Wang, Tanja Schultz, Alex Waibel "Comparison of Acoustic Model Adaptation Techniques on Non-native Speech",*IEEE International Conference on Acoustics, Speech and Signal Processing,* Vol 1, 2009

[9] R.Wallace, K.Thambiratnam, F. Scide, "Unsupervised Speaker Adaptation for Telephone Call Transcription",*IEEE International Conference on Acoustics , Speech and Signal Processing,* 2009

[10] Tara N. Sainath, Ron J. Weiss, Kevin W. Wilson,Bo Li, Arun Narayanan, Ehsan Variani, Michiel Bacchiani, Izhak Shafran, Andrew Senior, Kean Chin, Ananya Misra, and Chanwoo Kim, ``Multichannel Signal Processing With Deep Neural Networks for Automatic Speech Recognition'', *IEEE/ACM Transactions On Audio, Speech, And Language Processing,* Vol. 25, No. 5, May 2017

[11] Ali Haznedaroglu and Levent M. Arslan,"Language Model Adaptaion for Automatic Call Transcription", *IEEE International Conference on Acoustics, Speech and Signal Processing,*2014

[12] Seongjae Lee, Sunmee Kang, Hanseok KO, Jongseong Yoon, Minseok Keum, ``Dialogue Enabling Speech–to–Text User Assistive Agent with Auditory Perceptual Beamforming for Hearing-Impaired" *IEEE International Conference on Consumer Electronics (ICCE),* 2013

[13] Nuzhat Atiqua Nafis and Md. Safaet Hossain,`` Speech to Text Conversion in Real-time" *International Journal of Innovation and Scientific Research* Vol. 17 No. 2 Aug. 2015

[14] Rustam Shadiev, Barry Lee Reynolds, Yueh-Min Huang, Narzikul Shadiev, Wei Wang, Rai Laxmisha, Wanwisa Wannapipat,`` Applying speech-to-text recognition and computer-aided translation for supporting multi-lingual communications in cross-cultural learning project'', *IEEE 17th International Conference on Advanced Learning Technologies,* 2017

[15] Ricardo Sousa Rocha, Pedro Ferreira, Ines Dutra, Ricardo Correia, Rogerio Salvini, Elizabeth Burnside,`` A Speech-to-Text Interface for MammoClass'', *IEEE 29th International Symposium on Computer-Based Medical Systems,* 2016

[16] Chandra Mani Sharma, Alok Kumar Singh Kushwaha ,Rakesh Roshan , Rabins Porwal and Ashish Khare,`` Intelligent Video Object Classification Scheme using Offline Feature Extraction and Machine Learning based Approach", IJCSI *International Journal of Computer Science Issues,* Vol. 9, Issue 1, No 3, January 2012

[17] Vishwanath bigawan, Pinky kumari, Jordan pascual and Vijay bhaskar semwal, "Machine Learning Approach For Text And Document Mining", *arXIV preprint arXIV 1406.1580,*2014

[18] Stephan busemann, Svenschmeier and Roman.G.Arens "Message Classification In Call Centre", *Proceedings of Sixth Conference on*

Natural Language Processing, 158-164, 2009

[19] Parul Singla,Smriti Arya, Shailee L.Choudhary, ``Call Recording Mobile Application on Android", *International Journal of Scientific & Engineering Research,* Volume 7, Issue 1, January-2016

[20] Hongbo Liu, Member, Jie Yang, Member, Simon Sidhom, Yan Wang, Yingying Chen, and Fan Ye, ``Accurate WiFi Based Localization for Smartphones Using Peer Assistance'', IEEE *Transactions On Mobile Computing*, Vol. 13, No. 10, October 2014

[21] Jeong-Kyung Moon, Jin-Mook Kim, "Android API-based voice recording system in Emergency situations", *Springer Multimedia Tools and Applications*, Vol 76, Issue 17, 17971- 17984, September 2017

[22] Xavier Anguera Miro, Nicholas Evans, Corinne Fredouille, Gerald Friedland, and Oriol Vinyals, ``Speaker Diarization: A Review of Recent Research", *IEEE transactions on audio, speech, and language processing*, vol. 20, no. 2, February 2012

[23] Ji-Jian Wu, Jia-Yin Wang,`` A Voice-Enabled Android Roll Call System'', *IEEE International Conference on Consumer Electronics*, 2017

[24] Peng Luo, ``Design of Background Noise Suppression Circuit for Mobile Phones'', *Journal of Computer and Communications,* 2016

[25] Michel Vacher, Benjamin Lecouteux and Franc¸ois Portet, ``recognition of voice commands by multisource ASR and noise cancellation in a smart home

environment", 20 th European signal processing *Conference (EUSIPCO 2012) Bucharest, Romania,* August 27 - 31, 2012

[26] Zheng Yan, Christopher Niezreck, Louis N. Cattafesta , Diedrich O. Beusse, ``Background noise cancellation of manatee vocalizations using an adaptive line enhancer", *Acoustical Society of America,* 2006

[27] Andreas Schwarz, Christian Hofmann, Walter Kellermann, ``Combined Nonlinear Echo Cancellation and Residual Echo Suppression", *11th ITG Conference on Speech Communication Erlangen, Germany,* September 24 -26, 2014

[28] Stefan Gustafsson, Member, Rainer Martin, Peter Jax, and Peter Vary,`` A Psychoacoustic Approach to Combined Acoustic Echo Cancellation and Noise Reduction*", IEEE Transactions On Speech And Audio Processing*, Vol. 10, No. 5, July 2002

[29] Marvin R. Sambur, "Adaptive Noise Cancelling for Speech Signal*", IEEE Transactions on Acoustis, Speech and Signal Processing*, Vol

ASSP.26, No.5, 1978

[30] Urmila Shrawankar, Dr. Vilas Thakare. " Noise Estimation and Noise removal Techniques for Speech Recognition in Adverse Environment", *International Conference On Intelligent Information Processing*, 2010

[31] Prachi Khilari, Prof. Bhope V. P, ``Implementation of Speech to Text Conversion*" International Journal of Innovative Research in Science, Engineering and Technology* Vol. 4, Issue 7, July 2015

[32] S.A.R. Al-Haddad, Salina Abdul Samad, Aini Hussein, K. A. Ishak, A.A. Azid, R. Ghaffar, D. Ramli, M.R. Zainal, .K.A. Abdullah,

``Automatic Segmentation and Labeling for Continuous Number Recognition'' , *Proc. of the 6th WSEAS Int. Conf. on Signal Processing, Computational Geometry & Artificial Vision*, Elounda, Greece, August 21-23, 2006

[33] Kishori R. Ghule, R. R. Deshmukh, "Feature Extraction Techniques for Speech Recognition: Review" , *International Journal of Scientific & Engineering Research*, Vol. 6, Issue 5, 2013

[34] Sonia Sunny, David Peter S, K Poulose Jacob, ``Feature Extraction Methods based on Linear Predictive Coding and Wavelet Packet

Decomposition for Recognizing Spoken Words in Malayalam'',

International Conference on Advances in Computing and Communications, 2012

[35] Namrata Dave, "Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition" , *International Journal for Advance Resarch in Engineering and Technology*, Vol. 1, Issue VI, 2013

[36] Yogita H. Ghadage, Sushama D. Shelke , ``Speech to Text Conversion for Multilingual Languages, *International Conference on Communication and Signal Processing*", April 6-8, 2016

[37] Dr. G. Rasitha Banu ,VK Chitra, ``A Survey of Text Mining Concepts", *International Journel of Innovation in Engineering and Technology (IJIET)* vol.5 issue 2 April

2015

[38] B. Raghavendhar Reddy, E. Mahender,``Speech to Text Conversion using Android Platform", *International Journal of Engineering Research and Applications (IJERA)* Vol. 3, Issue 1, January -February 2013

[39] Talal Bin Amin, Iftekhar Mahmood, ``Speech Recognition Using Dynamic Time Warping'', *2nd International Conference on Advances in Space Technologies* Islamabad, Pakistan, 29th – 30th November, 2008

[40] Manjot Kaur, Lakhvir Garcha,``A Novel Approach for Speech Recognition using Vector Quantization through LBG Algorithm'',

International Journal of Engineering Science and Computing, July 2017

[41] Yoshua Bengio B, Réjean Ducharme, Pascal Vincent, Christian Jauvin, ``A Neural Probabilistic Language Model", *Journal of Machine Learning Research* 3 2003

[42] Grégoire Mesnil, Xiaodong He, Li Deng, and Yoshua Bengio,

``Investigation of Recurrent-Neural-Network Architectures and Learning Methods for Spoken Language Understanding" *Interspeech,* 25-29, August 2013

[43] Chee Peng Lim, Siew Chan Woo, Aun Sim Loh', and Rohaizan Osman', ``Speech Recognition Using Artificial Neural Networks'', *IEEE International Conference on Web Information Systems Enginering*, 419-423, Vol 1, 2000

[44] Su Myat Mon, Hla Myo Tun,`` Speech-To-Text Conversion (STT) System Using Hidden Markov Model (HMM)", *International Journal of Scientific & Technology Research* volume 4, issue 06, June 2015

[45] D.B. Pau, ``Speech Recognition Using Hidden Markov Models'', *The Liru:oln Laboratory Journal,* Volume 3, Number 1 (l990)

[46] Neha Sharma, Shipra Sardana, ``A Real Time Speech To Text Conversion System Using Bidirectional Kalman Filter In Matlab'', *Conference on Advances in Computing, Communications and Informatics (ICACCI), September,* 2016

[47] Morten Kolbæk, Dong Yu, Zheng-Hua Tan, and Jesper Jensen, ``Multitalker Speech Separation with Utterance-level Permutation Invariant Training of Deep Recurrent Neural Networks", *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* 2017

[48] Shivakumar K.M, Aravind K.G, Anoop T.V, DeepaGupta, ``Kannada Speech to Text Conversion Using CMU Sphinx'', *International Confrence on Inventive Computation Technologies,* 1-6, Vol 3, 2016

[49] Mehdi Allahyari, Seyedamin Pouriyeh, Mehdi Assefi,Saied Safaei, Elizabeth D.Trippe, JuanB.Gutierr, Krys Kochut, ``ABrief Survey of Text Mining:Classification,Clustering and Extraction Techniques", *KDD Bigdas, , Halifax, Canada,* August 2017

[50] Aurangzeb Khan, Baharum Baharudin, Lam Hong Lee, Khairullah khan, "A Review of Machine Learning

Algorithms for Text-Document Classification", *Journal of Advances in Information Technology,*Vol.1, No.1, February 2010

[51] Alper Kursat Uysal ,Yi Lu Murphey,``Sentiment classification: Feature selection based approaches versus deep learning" *IEEE International Conference on Computer and Information Technology,*2017

[52] Ms.Aradhana R. Patil, Ms.Amrita A. Manjrekar, ``An Innovative Approach to classifyand Retrieve Text Documents Using Feature Extraction and Hierarchical Clustering Based on Ontology'', *International Conference on Computing, Analytics and Security Trends (CAST),* 2016

[53] Anju Radhakrishnan , Vaidhehi V, ``Email Classification Using Machine Learning Algorithms", *International Journal of Engineering and Technology (IJET),* Vol 9 No 2 Apr-May 2017

[54] CAO Jian-fang WANG Hong-bin, ``Text Categorization Algorithms

Representations Based On Inductive Learning'', *International Conference on Information Management and Engineering,* 352-355, 2010

[55] Rakshana J. Shetty, Nithin Kumar Heraje, ``Recognition of Formatted Text using Machine Learning Technique", *American Journal of Intelligent Systems* 2017

[56] Nilesh Patil, Lokesh Patil, Gitanjali Wagh, Prof.Mis. Shital More,``

Candidate Recruitment System by Using Keyword Based Searching'', *International Research Journal of Engineering and Technology (IRJET)* Volume: 04 Issue: 03 | Mar -2017

[57] Herman Wandabwa,Defu Zhang, Korir Sammy, ``Text Categorization via Attribute Distance Weighted k-Nearest Neighbor Classification'', *International Conference on Information Technology*, 2016

[58] Ayman Mohamed Mostafa, ``An Evaluation of Sentiment Analysis and Classification Algorithms for Arabic Textual Data", *International Journal of Computer Applications (0975 – 8887)* Volume 158 – No 3, January 2017

[59] Er. Navjot Kaur, Er. Yadwinder Kaur, ``Object classification Techniques using Machine Learning Model", *International Journal of Computer Trends and Technology (IJCTT)* – Volume 18 Number 4 – Dec 2014

[60] Bastien Moysset, Christopher Kermorvant, Christian Wolf, ``Full-Page Text Recognition: *Learning Where to Start and When to Stop*", April 2017

[61] Mark HUGHES, Irene LI, Spyros KOTOULAS and Toyotaro

SUZUMURA, ``Medical Text Classification using Convolutional Neural Networks", *Informatics for Health:Connected Citizen-Led Wellness and Population Health, Stud Health Technol Inform.*235:246–50, 2017

[62] Wannaporn Teekeng Pornkid Unkaw , ``A New Hybrid Model of PSO and DE Algorithm for Data Classification", *IEEE/ACIS Internationcal Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing,*47-51, June 2017

[63] Gaurav Saxena, Manraj Singh Grover, Shampa Chakervarty, "Generating Word Clusters by Graph Clustering Based On Hearst Patterns'', *IEEE India International Conference on Information Processing*, 2016

[64] Yanmei Zhang, Tingpei Lei, Yan Wang, ``AService Recommendation Algorithm Based on Modeling of Implicit Demands'', *IEEE International Conference on Web Services*, 2016

[65] Kunhui Lin, Yating Chen, Xiang Li, Qingfeng Wu and Zhentuan Xu,``Friend *Recommendation Algorithm Based on Location-Based Social Networks", IEEE,* 2016

[66] Dongping Zhao, Jiapeng Xiu, Yu Bai, Zhengqiu Yang, "An Improved Item-Based Movie Recommendation Algorithm'', *IEEE International Conference on Cloud computing and Intelligence Systems,* 2016

[67] Greg Linden,Brent Smith,and Jeremy York, ``Amazon.com

Recommendations Item-to-Item Collaborative Filtering", *IEEE Internet Computing,* Vol 7, Issue 1, 76-80, January-February 2003

[68] Ibrahim Mashal, Osama Alsaryrah, and Tein-Yaw Chung, ``Analysis of Recommendation Algorithms for Internet of Things'', *Workshop on International Workshop on Mobile Edge Computing and IoT (MEIoT)* 2016

[69] Xingyi Zhang, Fuchen Duan, Lei Zhang, Fan Cheng, Yaochu Jin, Ke Tang,``Pattern Recommendation in Task-Oriented Applications: A Multi-Objective Perspective'', *IEEE Computational intelligenCe*

magazine, august 2017

[70] Erik van der Veen, Georgios Gousios, Andy Zaidman, ``Automatically Prioritizing Pull Requests'', *12th Working Conference on Mining Software Repositories*, 2015

[71] Dongmahn SEO, Suhyun KIM, Gyuwon SONG, and Seung-gil HONG, ``Speech-to-Text-based Life Log System for Smartphones*", IEEE International Conference on Consumer Electronics (ICCE)*, 2014