International Journal of Advanced Science and Technology
Vol. 29, No. 7s, (2020), pp. 3068-3076

Feature Extraction Algorithms for Speaker Recognition System and Fuzzy Logic

[1]Nilu Singh, [2]Nikhat Parveen, [3] Prabhash Chandra

[1]Assistant Professor, Department of Computer Applications, Babu Banarasi Das University, Lucknow, India.

[2]Associate Professor, Dept of CSE, Koneru Lakshmaiah Education Foundation, Guntur, A.P India.

[3]Associate Professor, Department of Computer Applications, Babu Banarasi Das University, Lucknow, India.

[1]nilu.chouhan@hotmail.com , [2] nikhat0891@gmail.com , [3] pathakprabhash2@gmail.com

Abstract

Speaker recognition is a technique used to identify a speaker by their voice. This technology used towards speech features for creating speaker's models and these models used for recognizing speakers. Speech feature extraction is performed by using some methods/algorithms and these known as feature extraction algorithms for speaker recognition. In this paper speech feature extraction algorithms are discussed and described with a comparative study. In addition, concept of fuzzy logic is discussed with their uses in speaker recognition.

Keywords: *Speaker Recognition, Speech features extraction, MFCC, LSF, PLP, DWT, LPC, LPCC, fuzzy logic, Type-2Fuzzy*

1. Introduction

Speaker Recognition (SR) is a technique which is used to recognize individuals by their voice. In this technical era it is one of the most important biometrics, since it is very easy to capture someone's voice by using any device which has a microphone or voice recorder. Basically speaker recognition is categorized into speaker verification and speaker identification.

Speaker verification means claimed speaker is either accepted or rejected while speaker identification means who from the registered speaker's is speaks. Speaker recognition can be text-dependent (fixed phrase or word spoken) or text-independent [1] [2]. Speaker recognition can be performed basically in three phases, these are acquiring speech signal for acoustic processing, feature extraction from speech signal, classification [3] [4] [6].

The aim of feature extraction techniques/algorithms for developing speaker/speech recognition system is to acquire an estimated value of short term spectral features for describing vocal tract characteristics [5]. Speech feature extraction is the process to convert a raw speech signal into an operational and compact representation which is more proficient to making adequate distinctions as compared to original signal [7].

Feature extraction is the first step of any speaker recognition system. Voice is unique due to differences in vocal cords and vocal tract shape, speaking style,

ISSN: 2005-4238 IJAST

International Journal of Advanced Science and Technology
Vol. 29, No. 7s, (2020), pp. 3068-3076

accent and some more acoustic aspects [8]. Speech signal contains many components which convey information about speaker. Vocal tract information expressed in the form of envelop of the short time power spectrum. The goal of speaker recognition system is to take those speech features which maximize between speaker variability and minimize within speaker variability [9] [10].

Process of speech feature extraction is known as front-end signal processing. This is achieved by altering speech waveform to a form of parametric depiction at a moderately lesser data rate for subsequent processing and analysis [12]. During front end processing processed speech signal transforms to a short form which is more reliable and judicial than original speech signal [11] [13]. Figure 1 show the basic steps involved in the automatic speaker recognition system.

In next section, feature extraction algorithms are described; section III represents a comparative table of different feature extraction methods. In section IV, concept of fuzzy logic in speaker recognition described. Finally conclusion has been made in section VI.

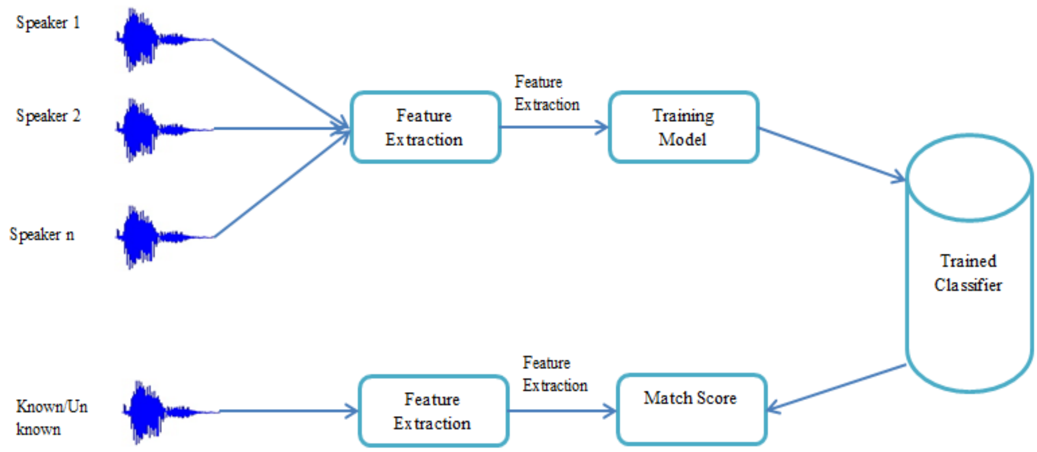


Fig.1 Steps involved in automatic speaker recognition system

2. Feature Extraction Algorithms

To develop a good automatic speaker recognition system, selection of robust feature extraction technique and an efficient modeling method are the core requirement of an SR system. Feature extraction is a process performs a decisive characteristic in speaker recognition. Generally, for acoustic measurement there are two methodologies used first is parametric method or temporal domain e.g. LP second is nonparametric frequency domain e.g. MFCC [14][15]. There are several speech signal processing algorithm was proposed by researchers such as LPC [16], LPCC, MFCC, PLP, DWT[17] etc. are commonly used [18-20]. Gaussian Mixture Model (GMM) is commonly used technique for Classification of speakers [22]. Following are some feature extraction methods/algorithms.

ISSN: 2005-4238 IJAST

2.1 Linear Prediction Coefficients (LPC)

It is widely used technique in many areas of speech signal processing. Coding and decoding of speech signal to save bandwidth and moderate data bit rate, many voice communication system used LPC [21]. This method is used for speech feature extraction and is used for both speech and speaker recognition. Here coefficient make out from speech signal frames, in addition as compare to modeling frame directly it provide better result if coefficient make out from speech frame sub-band decomposition. LPC performance is inadequate if speech signal is noisy. The main aim of to develop this method is to matching resonant structure (human vocal tract) that creates the equivalent sound [15]. The advantage of this technique is that its accuracy to estimate speech parameter and computation speed [21][23]. Figure 2 represents the steps involved in evaluating coefficient of LPC.

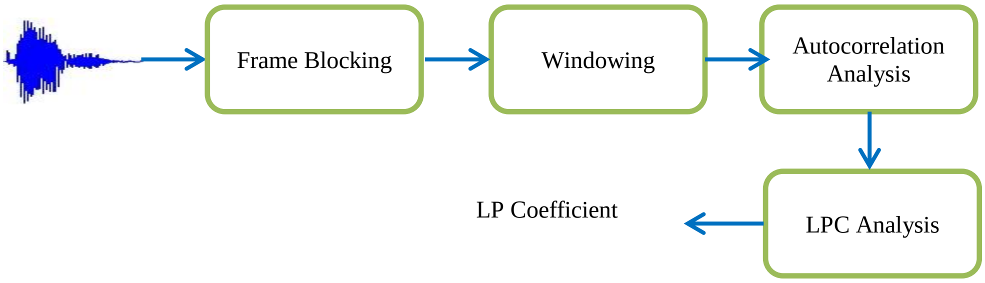


Fig.2 Process diagram of LPC

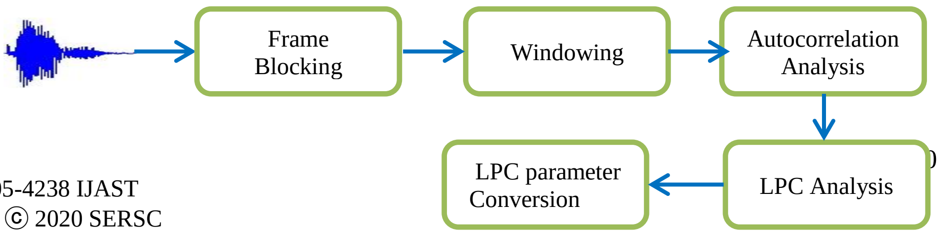
2.2 Linear Prediction Cepstral Coefficients (LPCC)

Generally phonetic statistics enclosed by spectral features since these are directly drawn out from spectra. Speech features derived from spectra are enrich with energy value of filter bank (linear), frequency component (equally emphasize) of any speech signal. To derive cepstral coefficient features linear prediction analysis or filter bank approach used. Using cepstral features, developed speaker recognition system produce high level accuracy result [7][17].

LPCC features also used for emotion recognition because it is consider that these feature having characteristics of vocal tract of a speaker. LPCC performance is inadequate if speech signal is noisy. To obtain cepstrum Linear Prediction (LP) analysis apply in a given speech signal. The concept behind using LP analysis is to estimate nth sample of a speech signal including linear combination of prior p samples [24][25]. Such as -

$$s(n) \approx a_1 s(n-1) + a_2 s(n-2) + a_3 s(n-3) + \cdots + a_p s(n-p)$$

Where $a1$, $a2$, $a3$, . . . are supposed constants of a speech analysis frame. Figure 3 represents a process diagram of LPCC.



ISSN: 2005-4238 IJAST

LPC Coefficient

Fig.3 Process diagram of LPCC

2.3 Mel-Frequency Cepstral Coefficients (MFCC)

This method is very effective in speaker as well as speech recognition system. It is based on human auditory perception system [26]. It is extensively used feature extraction technique in starting of 1990s for recognition since it is capable to represent compacted form of speech spectrum [27]. Popularity of MFCC is due to its capability to modeling the distinct frequency segment of speech signal, since it work on human's auditory perception. MFCC based on windowing the speech signal using DFT and fetching log of the magnitude after that deforming the frequencies on Mel scale by using inverse DCT [28-30].

This method is widely used for Automatic Speaker Recognition (ASR) due to its better performance in clean data and low computational complexity. Since degradation in performance is directly proportional to signal-to-noise ratio (SNR) hence performance degrades drastically if the speech signals noisy [15]. In figure 4 represented steps involved in evaluation of MFCC.

In MFCC, Mel-Spectrum is calculated by fourier transformed signal using band-pass filters in general called mel-filter bank. Here mel represent the frequency which is perceived by human ears. Normally mel scale is linearly physical frequency of tone of voice spacing below 1kHz and logarithmically above 1kHz [24]. Mel of a physical voice frequency is represented as

$$f_{mel} = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

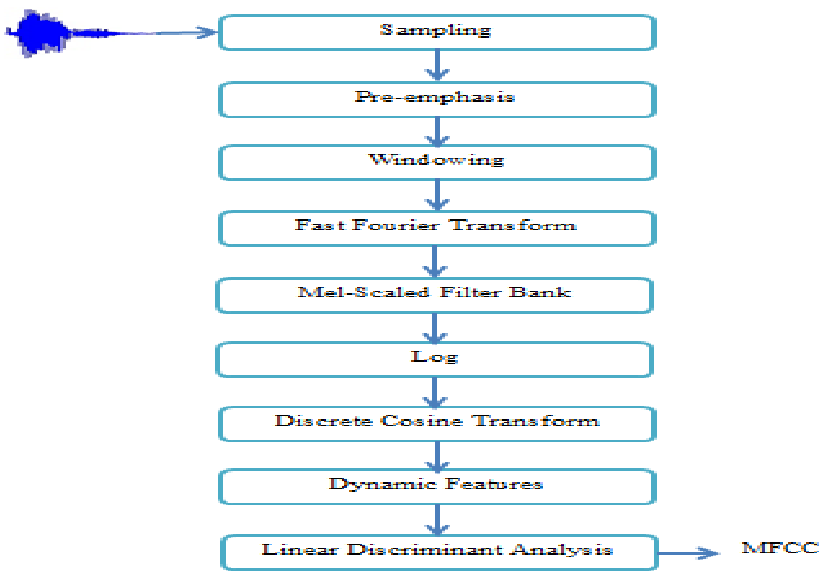


Fig.4 Process diagram of MFCC

ISSN: 2005-4238 IJAST

2.4 Perceptual linear prediction (PLP)

This method based on perceptual auditory characteristics. Using psychoacoustic process PLD features of a speech signal transform speech signal into a significant perceptual way [7]. In terms of human hearing, PLP analysis is more stable as compared to LP analysis. Perceptual linear predictive technique basically uses psychophysics of hearing to emanate auditory spectrum. These are: equal- loudness curve, critical band spectral resolution and intensity loudness power law [25]. After that autoregressive model is used to approximate the auditory spectrum. Figure 5 represents different stages involved in estimation of PLP.

To calculate perceptual linear prediction coefficient we can use the following recursion-

$$c_q = \begin{cases} ln(Q) & : q = 0 \\ -b_q + \frac{1}{m}\sum_{q=1}^{Q} -(m-q)b_q c_{m-q} & : q > 0 \end{cases}$$

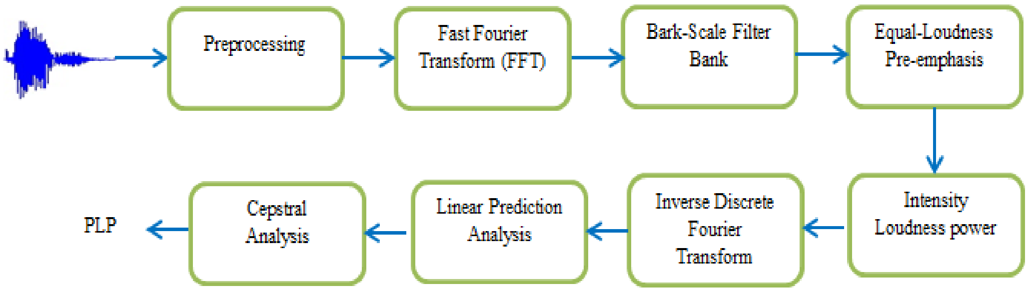


Fig.5 Process diagram of PLP

2.5 Line spectral frequencies (LSF)

Most of the developed speaker recognition systems use spectral features to create speaker model such as MFCC or PLP. It is projected by Itakura. Line spectral frequencies are good alternative of linear prediction coefficient. The studies show that LSF coefficient is significant and suitable for extracting speech feature characteristics.

Figure 6 represents different stages involved in LSF.It are shown in many studies that LSF has improved quantization properties as compared to other LP parametric. This method is also capable enough to decrease the bit-rate (25-30%) for transmission LP selective information without loss of the quality of speech signal. LSF basically represented as the predictor coefficient of inverse filter i.e. A(z). Initially A(z) break down into a duo of two auxiliary i.e. (p+1) like as follows-Generally LSF uses the frequency of zeros of P(z) and Q(z).

ISSN: 2005-4238 IJAST

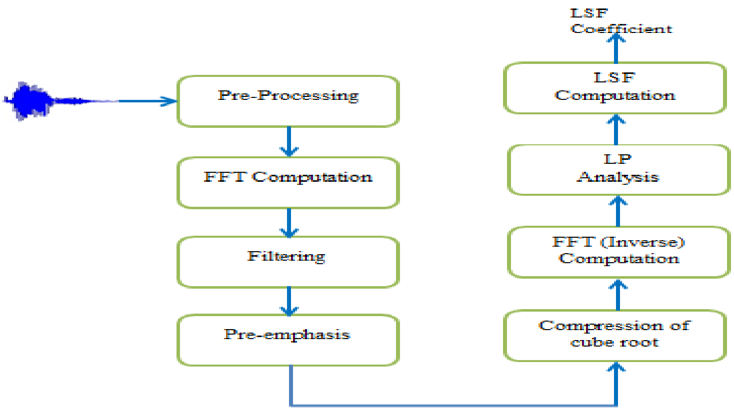


Fig.6 Different stages of Line Spectral Frequencies

2.6 Discrete wavelet transforms (DWT)

The main task of wavelet transform is to break-down speech signal into a set of functions known as wavelets. To examine the frequency spectrum WT used a variable window to increase the temporal resolution of the speech analysis. It is convenient for examining of non-stationary signals [7] [20]. In addition, it is a time frequency transform with the potential of multi-resolution. During feature extraction DWT apply on adaptive window size of speech features. The task of DWT is to break down the signals into sub-bands, so that easily features are differentiate from each sub-band. It is estimated by using the following formula.

$$\psi_{\tau,\alpha}(t) = \alpha^{-1/2}\psi(t - \tau/\alpha)$$

The DWT parameters enclose the information of speech signal with diverse frequency scales. Using wavelet transform technique non-stationary signals are represented with high efficiency [19]. This technique has capable enough to extract information in time and frequency domain from the transient signals [15]. Figure 7 shows the steps involved in extraction of DWT features.

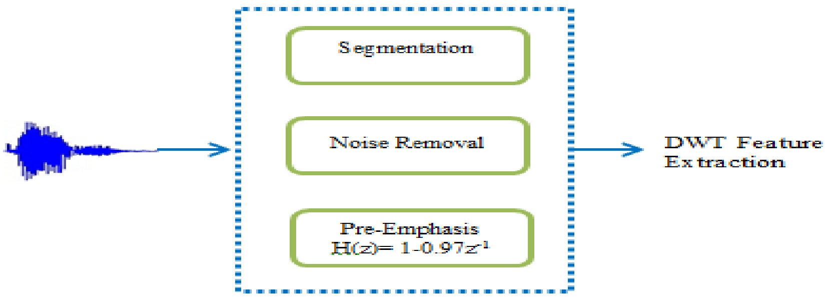


Fig.7 Process diagram of DWT

3. Comparison of Feature Extraction Technique

The Table 1 shown a comparative study of different feature extraction techniques based on some selected parameters, used in automatic speaker recognition system. The choice to select feature extraction algorithm is researcher interest and requirement. This table contains some crucial factors on the basis of these factors a

ISSN: 2005-4238 IJAST

comparison has been made. It may be helpful to those who somehow confused that which feature extraction algorithm is to be used for their work. Even though it is not said that which one is best but on the basis of some factors it may be decided.

4. Fuzzy Logic and Type-2 Fuzzy

In recent scenario concept of fuzzy theory is used in many research areas of computer science and engineering. Fuzzy logic comes into the existence in 1965 by Lotfi Zadeh as a solution for uncertainty. Fuzzy set theory used for speaker recognition (identification and verification) in many forms like Two Stage Fuzzy Decision Classifier (TSFDC), type-2 fuzzy, gender identification by fuzzy logic

.

In speaker recognition, uncertainty occurs during pattern matching due to variation in different parameters of speech signal. In such type of uncertainty problem which is related to parameters coped by fuzzy clustering algorithms. In addition also it is come in to notice that in past decade's researcher interest increases in Type-2 fuzzy sets due to potential of cope with uncertainties.

In speaker recognition Gaussian Mixture Model (GMM) is widely used for modeling the distribution of feature vectors of speaker‘s vocalization. But voice data is not always clean hence practical implementation facing uncertainties in the system parameters. In process of speaker verification/identification one main step is to use decision algorithm. Using fuzzy logic gender classification is cutting-edge research area where pitch is mainly used as a feature. The decision algorithm used characteristics of voices extracted by feature extraction techniques. There are some common decision making algorithm have used such as genetic algorithm, neural network and fuzzy logic. Figure 8 represents the steps involved estimation of fuzzy system decision.

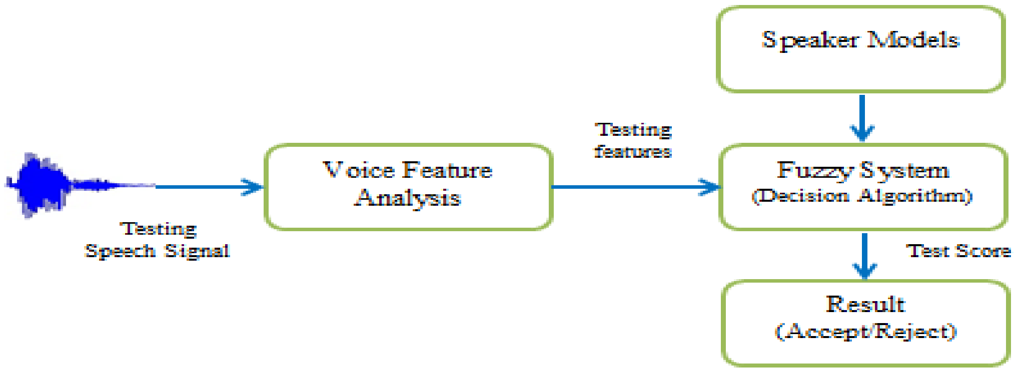


Fig.8. Process diagram of Fuzzy System

5. Findings

The selection of feature extraction technique is the core demand of developing a robust speaker recognition (verification/identification) system. There are many studies made by numerous researchers regarding significance of feature extraction methods/algorithms. Selection of a good, reliable feature extraction technique is a key factor to develop a more accurate speaker recognition system.

ISSN: 2005-4238 IJAST

6. Conclusion

This paper presented speaker recognition algorithms used for speech feature extraction. Also, discussion made about on the basis of requirement use the suitable feature extraction algorithm. Authors made discussion about some selected feature extraction techniques used for speech feature extraction for speaker recognition. Individual techniques have their importance and usefulness in different scenario. In addition, a comparative table has been made on the basis of different parameter to represent which feature extraction method select as per the system requirement.

References

[1] May, T., et al.,, "Noise-Robust Speaker Recognition Combining Missing Data Techniques and Universal Background Modeling", I E E E Transactions on Audio, Speech and Language Processing, 20(1), 2012, pp. 108-121.
[2] H. Hermansky and N. Morgan, "RASTA processing of speech", IEEE Transaction Speech Audio Process, vol. 2, no. 4, pp. 578–589, Oct. 1994.
[3] V. Mitra, et al.,, "Normalized amplitude modulation features for large vocabulary noise-robust speech recognition", in 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 2012, pp. 4117–4120.
[4] B. Bozilovic, et al., "Text-independent speaker recognition using two-dimensional information entropy," Journal of ELECTRICAL ENGINEERING, vol. 66, no. 3, July 2015, pp. 169-173.
[5] Nilu Singh, et al., "MFCC and Prosodic Feature Extraction Techniques: A Comparative Study", International Journal of Computer Applications", Published by Foundation of Computer Science, New York, USA, vol. 54(1), Sept. 2012, pp. 9-13.
[6] S. Velliangiri, P. Karthikeyan & V. Vinoth Kumar (2020) Detection of distributed denial of service attack in cloud computing using the optimization-based deep networks, Journal of Experimental & Theoretical Artificial Intelligence, DOI: 10.1080/0952813X.2020.1744196
[7] Praveen Sundar, P.V., Ranjith, D., Vinoth Kumar, V. et al. Low power area efficient adaptive FIR filter for hearing aids using distributed arithmetic architecture. *Int J Speech Technol* (2020). https://doi.org/10.1007/s10772-020-09686-y,
[8] Vinoth Kumar V, Karthikeyan T, Praveen Sundar P V, Magesh G, Balajee J.M. (2020). A Quantum Approach in LiFi Security using Quantum Key Distribution. International Journal of Advanced Science and Technology, 29(6s), 2345-2354.
[9] Tariq A. Hassan and Duraid T. Salim, "Arabic speech recognition using phase parameter", Journal of Physics: Conf. Series 1294 (2019) 042013, 2019, pp. 3-7.
[10] Umamaheswaran, S., Lakshmanan, R., Vinothkumar, V. et al. New and robust composite micro structure descriptor (CMSD) for CBIR. *International Journal of Speech Technology* (2019), doi:10.1007/s10772-019-09663-0
[11] Nilu Singh, et al., "Automatic Speaker Recognition: Current Approaches and Progress in Last Six Decades", Global Journal of Enterprise Information System (ISSN: 0975-1432, 2017), vol. 9, Issue-3, July-Sept. 2017, pp. 38-45.
[12] Karthikeyan, T., Sekaran, K., Ranjith, D., **Vinoth kumar, V**., Balajee, J.M. (2019) **"**Personalized Content Extraction and Text Classification Using Effective Web Scraping Techniques", *International Journal of Web Portals (IJWP),* 11(2), pp.41-52
[13] Boll SF, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans Acoustic Speech Signal Process, 10.1109/TASSP.1979.1163209, 1979, pp. 113-120.
[14] Vinoth Kumar, V., Arvind, K.S., Umamaheswaran, S., Suganya, K.S (2019), "Hierarchal Trust Certificate Distribution using Distributed CA in MANET", *International Journal of Innovative Technology and Exploring Engineering*, 8(10), pp. 2521-2524
[15] Nehe, N.S., Holambe, R.S, "DWT and LPC based feature extraction methods for isolated word recognition", Journal Audio Speech Music Proc. 2012, 7 (2012). https://doi.org/10.1186/1687-4722-2012-7 (Springer), 2012.

ISSN: 2005-4238 IJAST

[16] Maithili, K , Vinothkumar, V, Latha, P (2018). "Analyzing the security mechanisms to prevent unauthorized access in cloud and network security" *Journal of Computational and Theoretical Nanoscience*, Vol.15, pp.2059-2063.
[17] Makhoul J, "Linear prediction: a tutorial review", In: Proceedings of IEEE, vol 63, 1975, pp 561–580
[18] V.Vinoth Kumar, Ramamoorthy S (2017), "A Novel method of gateway selection to improve throughput performance in MANET", *Journal of Advanced Research in Dynamical and Control Systems*,9(Special Issue 16), pp. 420-432
[19] Gowdy JN, Tufekci Z, "Mel-scaled discrete wavelet coefficients for speech recognition", In Proc IEEE Inter Conf Acoustics, speech, and Signal Processing (ICASSP'00). Volume 3. Istanbul, Turkey; 2000, pp. 1351-1354.
[20] K Ravindranath, et.al., "Security key provided for Group Data Sharing in Cloud Computing "in the International Journal of Advanced Sciences and Technology (IJAST), ISSN: 2005-4238, November-2019.
[21] K Ravindranath, et.al., "An Advanced Secured Privacy Preserving Techniques for Cloud Using Numerical SQL Query's " in the International Journal of Advanced Sciences and Technology (IJAST), ISSN:2005-4238 , November-2019.
[22] P S RajaKumar, et.al., "Optimized and Efficient Computation of Big Data in Heterogeneous Internet of Things "in the International Journal of Engineering and Advanced Technology , ISSN:2249:8958,Volume -9, Issue-1, PP:6005-6010, October-2019.
[23] Arutselvan, et.al., "A Perspective of Probabilistic Misbehaviour Detection Scheme in Vehicular Ad-hoc Networks" in the International Journal of Innovative Technology and Exploring Engineering, SSN: 2278-3075, Volume - 8 Issue – 7,Page No:1098-1102, April 2019.
[24] G Sreeram, et.al., "Improving Cloud Data Storage Performance Based on Calculating Score using Data Transfer Rate Between the Internetwork Drives" in the International Journal of Engineering and Advanced Technology,ISSN: 2249 -8958, Volume-8 Issue-4,Page No: 1830-1835, April 2019.
[25] Dhilip Kumar V, Vinoth Kumar V, Kandar D (2018), "Data Transmission Between Dedicated Short-Range Communication and WiMAX for Efficient Vehicular Communication" *Journal of Computational and Theoretical Nanoscience*, Vol.15, No.8, pp.2649-2654
[26] K Sreenivasa Rao, et.al., "Detecting Fake Account On Social Media Using Machine Learning Algorithms" in the International journal of Control and Automation "2005-4297, Vol. 13, No. 1s, (2020), pp. 95-100, April 2020.
[27] G.Sreeram, et.al., "Efficiency and Stationing in Edge Computing "in the International Journal of Advanced Sciences and Technology (IJAST), ISSN: 2005-4238, Vol.29, 9s, (2020) pp.112-119.
[28] Kouser, R.R., Manikandan, T., Kumar, V.V (2018), "Heart disease prediction system using artificial neural network, radial basis function and case based reasoning" *Journal of Computational and Theoretical Nanoscience*, 15, pp. 2810-2817
[29] Shalini A, Jayasuruthi L, Vinoth Kumar V, "Voice Recognition Robot Control using Android Device" *Journal of Computational and Theoretical Nanoscience*, 15(6-7), pp. 2197-2201
[30] Jayasuruthi L,Shalini A,Vinoth Kumar V.,(2018) " Application of rough set theory in data mining market analysis using rough sets data explorer" *Journal of Computational and Theoretical Nanoscience*, 15(6-7), pp. 2126-2130

ISSN: 2005-4238 IJAST