Robust Feature Extraction Techniques in Speech Recognition: A Comparative Analysis

Isra Khan
PAF Karachi Institute of Economics & Technology

Rafi Ullah
PAF Karachi Institute of Economics & Technology

Shah Muhammad Emaduddin
PAF Karachi Institute of Economics & Technology

Robust Feature Extraction Techniques in Speech Recognition: A Comparative Analysis

Isra Khan[1], Ashhad Ullah Farooqui[2], Rafi Ullah[3], Shah Muhammad Emad[4]
College of Computing and Information Sciences,
Karachi Institute of Economics and Technology, Pakistan
[1]khanisra97@gmail.com [2]ashhadullah19@gmail.com [3]rafiafridi783@gmail.com [4]shahmuhammademad@gmail.com

Abstract – As the world is moving towards new era known as the era of 'Artificial intelligence' where many of things will be controlled automatically through many sources such as face and thumb lock like this we can control things through sound as the technology is evolving day by day so this technique is increasing rapidly but it is not explored that much. In this paper we are exploring sound and its feature extraction techniques through which we can extract features from various types of sound and can make them applicable as this paper presents a survey on feature extraction to comparative analysis with respect to properties such as noisy data, complexity, accuracy, extraction method it will be helpful to use which data set with which type of sound. Feature extractions process has a direct relation with any of the machine learning algorithm. If feature extracted is robust, the underlining machine learning algorithm used will be definitely accurate. This paper targeted only the comparative analysis of features used in literature for sound. In future, two or more features will be combined to enhance the impact of sound recognition systems.

Keywords- Sound Recognition, Feature Extraction in Sound Recognition, Sound Detection, Robust Feature in Sound Recognition and Detection, Robust Features in speech recognition

1. INTRODUCTION

Sound is the vibration that travels through air or any medium and these vibrations are audible when they reach an individual's ear. Sound is formed by the unbroken and consistent vibrations. The first ever device that can capture the sound waves was invented by Édouard-Léon Scott de Martinville, and was assembled by a gadget called a phonautograph in 1957. Phonautograph write out sound waves into a line that is drawn on paper but with these waves there are some features through which sound can be categorized in many classes or categories were extracted, let's take an example when we hear any kind of sound our brain start processing on it and categorize that sound like we can predict that this is the voice of a female without seeing the speaker because we know which frequency range belongs to which category of sound, but the major challenge is to extract the features, for this feature extracting purpose multiple techniques have been proposed such as MFCC, RASTA, LPCC, Cepstral Analysis, LPC and many others [1]. The majority of these proposed frameworks consolidate two handling stages. In first stage sound wave is captured and features are extracted from captured sound waves. The feature extraction process ordinarily includes a vast data reduction. Second stage does classification based on the extracted features and both of these stages are defined briefly below.
Many set of feature extraction are proposed earlier for audio classification [2, 3, 4]. Largest portion for audio classification has been covered by low-level signal extracting technique, second most important techniqueis Mel-frequency cesptral coefficient (MFCC) [5] and then rest of the features extracting techniques come.

All of the features extracted from sound are used audio classification and are very powerful in classifying the audio class but it gradually decrease when amount of classes increase. So, using which feature set with which amount of classes is an issue which can create further more issues if we select wrong feature set with respect to the problem description, result will help you with comparison done which will guide you when to use which set[6, 7].

Speech is that the most typical manner of communication between humans. Speech also carries the information related to the speaker. To recognize the speaker of the speech there are features exists in the speech signal. These extracted features will be useful in training of the model for speech recognition.

In audio processing, feature extraction is the backbone. The importance of feature extraction technique can never be ignored in speech recognition and processing systems [8]. But these features that are extracted must fulfill these criteria while doing speech recognition. These standards are [9]:

- Easy to measure extracted speech features
- Not be susceptible to mimicry
- Perfect in showing environment variation
- Stability over time

For feature extraction audio samples are collected and then converted to digital signals at a regular interval. At these voice samples noise reduction is performed so that the original audio sample can be find to perform feature extraction on it. For the speech recognition we extract the features from the digital signals which provide the acoustic properties of that specific digital dataset that is really useful for representing the speech signal.

These speech signals are slowly timed varying signals (quasi-stationary). When analyzed for a short time interval (5ms-100ms), the attributes seems to be relatively stationary. However if sound/vocal features are modified for a specified time interval, it reflects the different values of spoken audio features. The information of audio signal can be categorized by using short term amplitude spectrum of the audio wave form. This techniques known as phonemes helps in the extraction of sound features of short term amplitude spectrum from audio signals called phonemes [10].

Rest of the paper is divided as follow; Section I is about the literature review or Related Work, Section II is the detail explanation of different features that can be extracted from sound, Section IV is Result section that is detail comparative analysis of different features extraction technique, Section V is the concluding the topic and Section VI is the future potential area.

2. Related Work

Authors of [11] focused on the comparative analysis of widely used feature extraction techniques related to speech recognition and in the end of the research has conclude that the PLP is extracted on the conception of logarithmically spaced filter bank, combined with the conception of human hearing system and has improved results than LPC.

According to paper [29], author has extracted MFCC feature and de-noise the audio sample and also enhanced the MFCC feature by calculating the delta energy for the coefficient.

Authors has extracted MFCC feature for the speech emotion detection discussed in detail in [30]. MFCC feature is extracted and worked very efficiently and train the model for the detecting of speech detection emotion.

Isolated speech recognition by using the MFCC and Dynamic Time Wrapping (DTW) was focused by the authors of [31]. In this research features for the isolated speech recognition were extracted by using the MFCC.

In research [14], authors has identified and focused on the problem of optimizing the acoustic features set by Ant Colony Optimization for the Automatic speech recognition. Speech signal is considered as input in this research and feature extraction is performed over this signal using MFCC extraction method, total 39 coefficients are extracted in this research by using MFCC.

Comparative analysis of speech recognition has done in paper [33]. These analysis was performed on noisy conditions on the widely used feature extraction techniques named MFCC,LPCC,PLP, RASTA-PLP and HMM and has analyzed that PLP distinctly gave the maximum percentage of recognition and the grouping of LPCC, PLP and RASTA provided the output as third maximum recognition percentage.

In paper [34], writers have worked on the change detection in multi-dimensional unlabeled data in which features were extracted by using the PCA feature extraction technique.

According to the authors of [35], they focused on the PCA drawbacks which are high computational cost, extensive memory utilization and low adequacy in handling expansive dimensional datasets, so author has proposed a new technique named as Folded-PCA. By using this new proposed technique these drawbacks can be resolve.

Drawbacks of PCA were discussed in paper [36]. These drawbacks are: computational cost, extensive memory utilization and low adequacy in handling expansive dimensional datasets, so they analyzed two variation of the PCA technique SPCA and Seg-PCA. These variations can be helpful to reduce the drawbacks of PCA.

Authors in [20] have done the survey over the feature extraction technique and conclude that the LPC is vector dimension and has high computational cost and also reduce accuracy and their window size which is not good for non-stationary speech signals such as speech signal.

In [38], writers has proposed the new technique for the noisy speech recognition based on auditory filter modeling-based feature extraction and gives the result that LPC is less efficient in this manner in comparison with PLPaGc.

Comparative analysis for the speech recognition specific for Hindi language words, and has analyzed that LPCC gives less recognition rate for isolate, paired and hybrid words as compared to MFCC has performed in [39].

A new recognition system was proposed in [40]. This system uses the acoustic waves generated by the construction equipment; this will be very helpful to avoid external damages. Feature extraction for the recognition system was done by combining LPCC and SVM.

RASTA feature extraction technique in combination with TANDEM was used by the authors of [41]. The authors stated that this technique is an efficient way to represent the message-information in the speech signal.

3. Feature Extraction Techniques

Various Features Extraction techniques have been observed in literature used for sound recognition and sound detection. Each one has its own advantages and disadvantages depending upon the environment i.e. nature of problem. For example features extraction used in sounds related to school cafe will be having different impact on sound of vehicles. Some of the features extraction techniques that are observed during our research are:

- Mel-frequency Cepstral Coefficients (MFCC)
- Perceptual Linear Predictive (PLP)

- Relative Spectral Processing (RASTA)
- Linear Prediction Cepstral Coefficient (LPCC)
- Principle Component Analysis (PCA)
- Linear Discriminant Analysis (LDA)
- Wavelet
- Dynamic Time Warping (DTW)
- Combined LPC and MFCC
- Kernel based feature extraction
- Independent Component Analysis (ICA)
- Integrated phoneme subspace method
- Probabilistic Linear Discriminant Analysis (PLDA)
- Linear Prediction Coefficient
- Discrete Wavelet Transformation (DWT)
- Wavelet Packet Decomposition (WPD)
- Gammatone Frequency Cepstral Coefficient (GFCC)
- Gaussian Mixture Model (GMM)

This paper targeted only six features extraction technique (MFCC, PLP, PCA, LPC, LPCC and RASTA) to compare on the basis of several parameters such as Impact in presence of noise i.e. noisy data, Complexity (in case of features extraction and computation), Accuracy and Feature Extraction Method.

I. Mel-frequency Cepstral Coefficients

MFCC is one of the most important techniques used to extract the feature from speech signal [11] that is actually based over the human's ear scale bandwidth. It uses the low and the high frequencies, measured in Hertz (Hz) to get the speech signal. MFCCs considered as frequency domain features that are more accurate in comparison with time domain features [11]. These signals are then divided into the audio frames to calculate the MFCC. Assume each frame of audio signal contains the N samples and considers the next and the previous frames of the audio signal is separated by M samples where M < N then all audio frames are multiplied by a Hamming window. The hamming window [16] value can be calculated using this equation 1.

$$W(n) = 0.54 - 0.46\cos(2\pi n/N-1) \qquad (1)$$

Then speech signal is transformed to frequency domain from time domain by utilizing its Discrete Fourier Transform. The Mel frequency scale [17] is consider as linear frequency having spacing less than 1000 Hz and a logarithmic spacing more than 1000Hz .As a reference point ,a pitch of a 1 KHz tone ,40 dB above the threshold perceptual hearing, is defined as 1000 mels. So, to find the mels for a specific given frequency f in Hz we can use this equation 2.

$$Mel(f) = 2595*\log10(1 + f/700) \qquad (2)$$

The MFCC features correspond to the total power of the log in a critical band around the center frequencies. Finally, for the calculation of cepstral coefficients, the Inverse Discrete Fourier Transformer is applied; finally calculate the DCT of the output from the filter-bank. The resultant value is the actual Mel-Frequency Cepstral Coefficient.

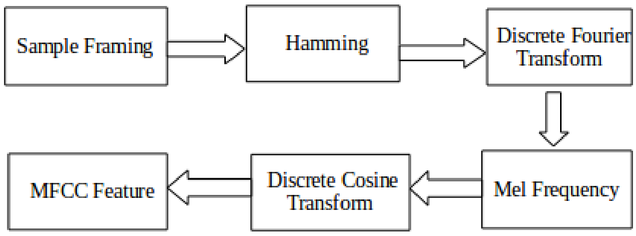

Figure 1 : MFCC features Extraction Technique

II. Perceptual Linear Predictive

The PLP model aims at human vocalizations based on the concept of hearing psychophysics and then more precisely in the process of extracting features. PLP increases the rate of speech recognition and also eliminates irrelevant speech information [18]. PLP technique is quite similar to LPC but differs from MFCC. PLP mainly consists of three steps. First one is for critical band analysis. Second is for equal loudness and the third one is for intensity-loudness and power-law relation. PLP carries out spectral analysis with frame of N samples with N band filters on the speech vector. For the experiments, 256 window sizes and 24 filter banks are used. The PLP filters are then produced with pre-emphasis and scale of bark. Next step is the estimation of power spectrum with the power law [18]. Now computed PLP spectrum is forwarded to LP analysis with the frequencies. Atlast LP analysis is performed along FFT and then final values are observed by calculating the inverse of FFT.

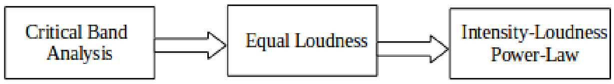

Figure 2: PLP features Extraction Technique

III. Linear Prediction Coefficient

LPC is actually works on the prediction. In samples of speech signal we can predict the nth samples, which can be represented by summarizing the previous samples of the target signals (k). The inverse filter production should be carried out to match the formants region of the speech samples [19]. The LPC process is therefore the application of these filters in the samples [20]. The main idea of LPC is to approximate the current (n) acoustic sample s(n) with the previous samples s(p).

$$s(n) \approx a_1(n-1) + a_2(n-2) \pm \ldots + a_p(n-p) \qquad (3)$$

Then LPC is obtained using the Levinson-Durbin recursive algorithm [20]:

$$H(z)=\frac{1}{1-\sum_{k-1}^{p} a_k z^{-k}} \quad (4)$$

H(z) reflects the propagation path of the acoustic signal. Let c (n) be the impulse response [20]:

$$\widehat{H}(z)=lnH(z)=\sum_{n=0}^{\infty} c(n) z^{-k} \quad (5)$$

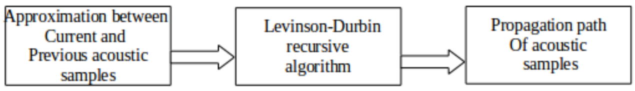

Figure 3: LPC features Extraction Technique

IV. Linear Prediction Cepstral Coefficient

Liner Prediction Cepstral Coefficient is an enhanced version of LPC method. The representation of linear predictive coefficients in cepstrum domain can be reflected as new coefficients known as linear predictive cepstral coefficients [21]. The value of LPCC coefficient can be computed by using LPC equations which are as follows.

$$c_1=a_1 \quad (6)$$

$$c_n=a_n+\sum_{k-1}^{n-1} \frac{k}{n} c_k a_{n-k} \; 1<n\le p \quad (7)$$

$$c_n=a_n+\sum_{k-1}^{n-1} \frac{k}{n} c_k a_{n-k} \; n>p \quad (8)$$

Where $c_1, c_2, \ldots, c_n$ are the LPCC.

V. Principle Component Analysis

PCA is thought a Principle part Analysis – this is often a statistical analytical tool that's used to explore kind and cluster information. PCA take an over-sized variety of correlate (interrelated) variables and rework this information into a smaller variety of unrelated variables (principal components) whereas holding largest quantity of variation, so creating it easier to work the information and build predictions. PCA could be a method of distinguishing patterns in information, and expressing the information in such some way on highlight their similarities and variations. Since a pattern in information is hard to seek out in information of high dimension, wherever the posh of graphical illustration isn't offered, PCA could be a powerful tool for analyzing information [10].

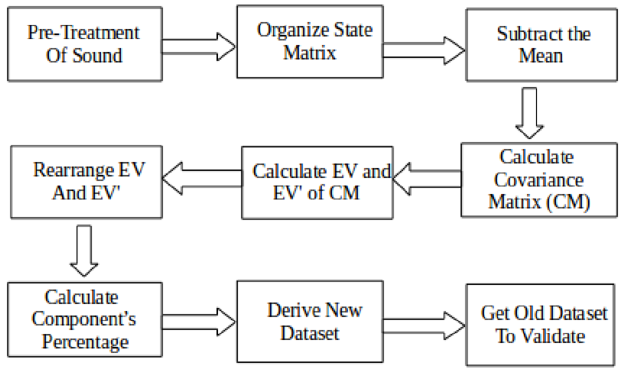

Figure 4 : PCA features Extraction Technique

Where EV is Eigen Vector and EV' is Eigen Value.

VI. Relative Spectral Processing

RASTA is method of extracting the relevant information from a sound or any speech signal and the main objective of this technique is to eliminate the robustness of speech recognition in noise or in the real time environments [16] and it is usually done by using time trajectories of band pass filter of logarithmic speech value, in fact it is the extension of the original method by combining additive noise and convolution noise [15].

RASTA is a voice improvement based on linear filtering of the short-term power spectrum of the noisy audio signal, as shown in Figure 5. The input speech signal spectral values are compressed by a nonlinear compression rule (a=2/3) before filtering and expanded after filtering (b=3/2) [16].
Output of each filter is given as,

$$S_i(K)=\sum_{j=-M}^{M} W_i(j) Y_i(k) \quad (9)$$

Where $S_i(k)$ is a clean speech estimate , $Y_i(k)$ is the noisy audio spectrum $W_i(j)$ is the filter weights and M is the filter order.
These values can be set according to the required processing or the corresponding set of problem.

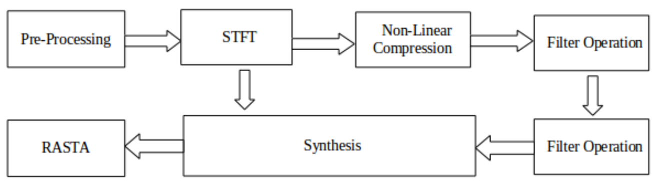

Figure 5 : RASTA features Extraction Technique

4. Results

Details comparison observed during this survey are following that will help researchers and practitioner to know the insights of different feature extraction techniques used in sound recognition and detection problems.

TABLE 1: COMPARATIVE ANALYSIS OF FEATURES EXTRACTION TECHNIQUES USED IN SOUND RECOGNITION AND DETECTION

Features Extraction	**Noisy Data**	**Complexity**	**Accuracy**
MFCC	Poor result on noisy data [27]	Less Complex and High performance rate	92% [24]
PLP	poor result on noisy data due to spectral balance of formant	Slightly Complex	Better Performance than LPCC and MFCC [25]
PCA	Doesn’t work well on noisy data as it does not reduce noise completely.	Slightly Complex and High Performance Rate	54.66% [8]
LPC	Not good for noisy data [27].	Less complex [27].	Good Accuracy, reliability and robustness [24]
LPCC	Shows poor result on highly noised data [27].	Simple and good performance [27]	Accuracy is 88% [26]
RASTA	Works good on noisy data as it enhances data by removing noisy distortions [27].	Slightly Complex	A robust technique. Low modulation frequencies are captured [24].

TABLE 2: COMPARATIVE ANALYSIS OF FEATURES EXTRACTION TECHNIQUES USED IN SOUND RECOGNITION AND DETECTION

Features Extraction	**Extraction Method**	**Final Comments**
MFCC	Dynamic method [27].	Mostly used where human ear bandwidth scale exists.
PLP	Combines the linear prediction analysis and spectral analysis [9]	Increases the recognition rate and also removes noise.
PCA	Non-Linear method [27].	Eigen vector based. Reduce Components / Dimensions of Features
LPC	A static method [29].	Used for extraction at lower rate. It can be used in sound recognition of abnormal sounds
LPCC	Use Autocorrelation analysis [27].	Used in cepstral domain.
RASTA	Non-Linear Compression [16].	Highly recommended in domain where there is noise, it will extract good features in noisy data

5. CONCLUSION

In this paper we have discussed some widely used feature extraction techniques in the domain of speech recognition. The motivation for doing this comparative analysis is because there are many feature extraction techniques are available and very few of them are really helpful. This paper will guide the researchers regarding the feature extraction technique and will also help them to differentiate between them.

6. FUTURE WORK

Novel and Robust features can also be extracted by combining many of the existing features to enhance the capability of sound detection and recognition systems. Developing a system that will record complete meeting conversation in a dialogue form, sentence spoken by each person against their name (if known), otherwise a separate line by some person “i”. This system will reduce time of recording meeting or writing manual points where some points may be skipped or interpreted wrong.

REFERENCES

[1] McKinney, Martin, and Jeroen Breebaart. "Features for audio and music classification." (2003).

[2] Davis, Steven, and Paul Mermelstein. "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences." *IEEE transactions on acoustics, speech, and signal processing* 28, no. 4 (1980): 357-366.

[3] Wold, Erling, Thom Blum, Douglas Keislar, and James Wheaten. "Content-based classification, search, and retrieval of audio." *IEEE multimedia* 3, no. 3 (1996): 27-36.

[4] E. Scheirer and M. Slaney. Construction and evaluation of a robust multifeature speech/music discriminator. In Proc. ICASSP, pages 1331–1334, Munich, Germany, 1997.

[5] H. Hermansky and N. Malayath. Spectral basis functions from discriminant analysis. In International Conference on Spoken Language Processing, 1998

[6] M. Zhang, K. Tan, and M. H. Er. Three-dimensional sound synthesis based on head-related transfer functions. J. Audio. Eng. Soc., 146:836–844, 1998.

[7] T. Zhang and C. C. J. Kuo. Audio content analysis for online audiovisual data segmentation and classification. IEEE Transactions on speech and audio processing, 2001.

[8] Prasad, K.S., Ramaiah, G.K. and Manjunatha, M.B., 2017. Speech features extraction techniques for robust emotional speech analysis/recognition. *Indian Journal of Science and Technology*.

[9] Vimala, C. and Radha, V., 2014. Suitable feature extraction and speech recognition technique for isolated tamil spoken words. *International Journal of Computer Science and Information Technologies (IJCSIT)*, *5*(1), pp.378-383.

[10] Shrawankar, U. and Thakare, V.M., 2013. Techniques for feature extraction in speech recognition system: A comparative study. *arXiv preprint arXiv:1305.1145*.

[11] Dave, N., 2013. Feature extraction methods LPC, PLP and MFCC in speech recognition. *International journal for advance research in engineering and technology*, *1*(6), pp.1-4.

[12] Barchiesi, D., Giannoulis, D., Stowell, D. and Plumbley, M.D., 2015. Acoustic scene classification: Classifying environments from the sounds they produce. *IEEE Signal Processing Magazine*, *32*(3), pp.16-34.

[13] Desai, N., Dhameliya, K. and Desai, V., 2013. Feature extraction and classification techniques for speech recognition: A review. *International Journal of Emerging Technology and Advanced Engineering*, *3*(12), pp.367-371.

[14] Këpuska, V.Z. and Elharati, H.A., 2015. Robust speech recognition system using conventional and hybrid features of mfcc, lpcc, plp, rasta-plp and hidden markov model classifier in noisy conditions. *Journal of Computer and Communications*, *3*(06), p.1.

[15] H. Hermansky and N. Morgan, "RASTA processing of speech," in *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 4, pp. 578-589, Oct. 1994.

[16] Kurzekar, P.K., Deshmukh, R.R., Waghmare, V.B. and Shrishrimal, P.P., 2014. A comparative study of feature extraction techniques for speech recognition system. *International Journal of Innovative Research in Science, Engineering and Technology*, *3*(12), pp.18006-18016.

[17] Tiwari, V., 2010. MFCC and its applications in speaker recognition. *International journal on emerging technologies*, *1*(1), pp.19-22.

[18] Chandra, E., and K. Manikandan M. Sivasankar. "A proportional study on feature extraction method in automatic speech recognition system." (2014).

[19] Narang, S. and Gupta, M.D., 2015. Speech Feature Extraction Techniques: A Review. *International Journal of Computer Science and Mobile Computing*, *4*(3), pp.107-114.

[20] Yang, S., Cao, J. and Wang, J., 2015, July. Acoustics recognition of construction equipments based on LPCC features and SVM. In *Control Conference (CCC), 2015 34th Chinese* (pp. 3987-3991). IEEE.

[21] Kaur, K. and Jain, N., 2015. Feature Extraction and Classification for Automatic Speaker Recognition System-A Review. *International Journal of Advanced Research in Computer Science and Software Engineering*, *5*.

[22] Gupta, D., Bansal, P. and Choudhary, K., 2018. The state of the art of feature extraction techniques in speech recognition. In *Speech and Language Processing for Human-Machine Communications* (pp. 195-207). Springer, Singapore.

[23] Chaudhary, G., Srivastava, S. and Bhardwaj, S., 2017. Feature Extraction Methods for Speaker Recognition: A Review. *International Journal of Pattern Recognition and Artificial Intelligence*, *31*(12), p.1750041.

[24] Gupta, H. and Gupta, D., 2016, January. LPC and LPCC method of feature extraction in Speech Recognition System. In *Cloud System and Big Data Engineering (Confluence), 2016 6th International Conference* (pp. 498-502). IEEE.

[25] Khara, S., Singh, S. and Vir, D., 2018, April. A Comparative Study of the Techniques for Feature Extraction and Classification in Stuttering. In *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)* (pp. 887-893). IEEE.

[26] Kianisarkaleh, A. and Ghassemian, H., 2016. Spatial-spectral locality preserving projection for hyperspectral image classification with limited training samples. *International journal of remote sensing*, *37*(21), pp.5045-5059.

[27] Singh, P.P. and Rani, P., 2014. An approach to extract feature using mfcc. *IOSR Journal of Engineering*, *4*(8), pp.21-25.

[28] Kaur, J. and Sharma, A., 2014. Emotion detection independent of user using mfcc feature extraction. *International Journal of Advanced Research in Computer Science and Software Engineering*, *4*(6).

[29] Dhingra, S.D., Nijhawan, G. and Pandit, P., 2013. Isolated speech recognition using MFCC and DTW. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, *2*(8), pp.4085-4092.

[30] Poonkuzhali, C., Karthiprakash, R., Valarmathy, S. and Kalamani, M., 2013. An approach to feature selection algorithm based on ant colony optimization for automatic speech recognition. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, *2*(11), pp.5671-5678.

[31] Kuncheva, L.I. and Faithfull, W.J., 2014. PCA feature extraction for change detection in multidimensional unlabeled data. *IEEE transactions on neural networks and learning systems*, *25*(1), pp.69-80.

[32] Zabalza, J., Ren, J., Yang, M., Zhang, Y., Wang, J., Marshall, S. and Han, J., 2014. Novel Folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *93*, pp.112-122.

[33] Ren, J., Zabalza, J., Marshall, S. and Zheng, J., 2014. Effective feature extraction and data reduction in remote sensing using hyperspectral imaging [applications corner]. *IEEE Signal Processing Magazine*, *31*(4), pp.149-154.

[34] Hibare, R. and Vibhute, A., 2014. Feature extraction techniques in speech processing: a survey. *International Journal of Computer Applications*, *107*(5).

[35] Zouhir, Y. and Ouni, K., 2014. A bio-inspired feature extraction for robust speech recognition. *SpringerPlus*, *3*(1), p.651.

[36] Gulzar, T., Singh, A. and Sharma, S., 2014. Comparative analysis of LPCC, MFCC and BFCC for the recognition of Hindi words using artificial neural networks. *International Journal of Computer Applications*, *101*(12), pp.22-27.

[37] Hermansky, H. and Fousek, P., 2005. *Multi-resolution RASTA filtering for TANDEM-based ASR* (No. REP_WORK). IDIAP.