iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** X **Month of publication:** October 2023

DOI: https://doi.org/10.22214/ijraset.2023.55454

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue X Oct 2023- Available at www.ijraset.com

Speech Enhancement Based on Approximate Message Passing

K. Yeshwanth[1], Dr. K. Santhi Sree[2]
[1]*Student, M.Tech (Data Sciences), School of Information Technology, JNTU Hyderabad, Hyderabad, India*
[2]*Professor of CSE & Head, School of Information Technology, JNTU Hyderabad, Hyderabad, India*

Abstract: ***A novel speech improvement technique based on the approximate message passing (AMP) is adopted to get beyond the drawbacks of traditional speech enhancement methods, such as inaccurate voice activity detector (VAD) and noise estimation. To eliminate or muffle the noise from the distorted speech, AMP takes advantage of the difference between speech and noise sparsity. The AMP method is used to effectively rebuild clean speech for speech augmentation. More specifically, the prior probability distribution of the speech sparsity coefficient is represented by a Gaussian model. The expectation maximization (EM) technique provides excellent learning of the hyper-parameters of the prior model. We use the k-nearest neighbor (k-NN) approach to learn the sparsity while taking into account the correlation between the speech coefficients in neighboring frames. Additionally, computational simulations are used to verify the proposed algorithm, which outperforms the three methods which are Kalman filter, principal component analysis (PCA), independent component analysis (ICA) under a variety of signal to noise ratios and compression ratios (SNRs).***

Keywords: Kalman filter; Digital audio effect; digital expander; Independent component analysis; principal component analysis.

I. INTRODUCTION

Speech plays a vital part in our diurnal communication and mortal- machine interfacing. Thus, product and perception of speech have come an intriguing part of the exploration for decades. But the quality and intelligibility of the speech are significantly degraded by background noise, which affects the capability to understand others' speech, causes crimes in mortal Machine Interfacing, etc. In this digital world, it's really hard for any signal in a real- time terrain to escape from noise. This hits us hard when it comes to delivering a communication from one place to another, and there's a need for drawing up or enhancing the communication signal but, at the same time, not giving up any intelligibility of the communication (content, not just clarity). Since speech dispatches have been the mode of communication far and wide, speech improvement is needed whenever the signal comes in contact with the real- time terrain. Modelling the mortal speech product process helps in enhancing speech. But, as speech is a largely nonstationary signal, it is not easy to model the mortal speech product process. Though speech is a largely nonstationary signal, it's stationary for a veritably short period. Grounded on this fact, Classical speech improvement ways are considered for speech member models for a short time, but these short time models don't include the goods of the noise as noise has long- term characteristics. The AR model is also known to be good for representing unspoken speech. Still, it’s unhappy for raised speech since it’s frequently periodic. This has motivated us to look into speech models that can satisfactorily describe both raised and unspoken speech and allow for the exploitation of long- term noise characteristics. Speech improvement is an area of speech processing where the thing is to ameliorate the intelligibility and niceness of a speech signal. The most common primary ideal of numerous Speech Enhancement algorithms is to ameliorate the perceptual quality of rooting speech signals from noisy speech. Noise estimation is the major element in speech improvement ways because better noise estimation gives a high quality of speech birth. Till now, removing noise from noisy speech has been a grueling issue because spectral parcels of nonstationary noise are veritably delicate to estimate and prognosticate. Noise estimation is a careful issue in speech improvement algorithms since if the noise power is further than speech power, also that speech content may be removed due to treating that as noise. Numerous operations like teleconferencing systems, speech recognition- grounded security bias, biomedical signal processing, hearing aids, ATMs, and computers, Speech improvement is a hot exploration area in signal processing. It remains a grueling issue because, in utmost cases, only noisy speech is available. Over the once times, experimenters have developed different types of effective algorithms to ameliorate noisy speech indeed though it still poses a challenge to the experimenters because of characteristics of noise signals vary dramatically over time and operation to operation. Numerous speech improvement ways have been proposed using a filtering approach by experimenters last ten times, similar as the spectral deduction system, wiener filtering, Kalman sludge system, etc. Spectral deduction is used for enhancing speech degraded by cumulative stationary background noise.

©IJRASET: All Rights are Reserved

Still, it's affected by musical noise and doesn't remove noise during the silence period. The Wiener sludge- grounded speech improvement system recovers the original signal by minimizing Mean Square Error (MSE) between the clean speech and the estimated signal. Spectral and wiener sludge- grounded speech improvement algorithms bear the characteristics of clean speech. But in real- time, clean speech may not be available in all cases.

From the literature study, we set up that some of the ways have been proposed to enhance speech. Using the harmonious structure of speech signals, a speech signal is recovered from a noisy speech signal; the sinusoidal model is espoused in the MMSE estimator to enhance the speech introduced by Ephraim. At first, The advantage of using the Kalman sludge for speech improvement using estimation of speech signal parameters from clean speech before it's corrupted by white noise is proposed. And further extended to the arbitrary and colored noises. In these styles, a trade- off should be maintained between SNR and intelligibility. Latterly, numerous changes were made to the Kalman sludge for better enhancement; it didn't meet the prospects or complexity. In this paper, with lower complexity and better performance, a new adaptive Kalman sludge- grounded system with the combination of a nonlinear digital sludge called a digital expander is proposed to recover the speech signal from noisy speech. The cumulative noise is modelled as the AR process grounded on the Kalman filtering algorithm's direct vaticination measure estimation (LPC). In addition to measure estimation, this paper answered the problem of de-noising the arbitrary and colored noises. We considered a supposition that the colored noise is also an autoregressive process. So we estimated its AR portions and dissonances by direct vaticination estimation also. In this paper, to overcome above stated problem, a new adaptive Kalman sludge- grounded system with pre-processing of a digital audio effecting fashion called a digital expander is proposed to recover the speech signal from a sequence (frame) of noisy speech signals, and the cumulative noise is modelled as the AR process. This time- varying bus-accumulative (AR) speech model parameter estimation is grounded on direct vaticination measure estimation (LPC). In addition to measure estimation, this paper answered the problem of de-noising the colored noise. We assumed that the noise is also an autoregressive process. So we estimated its AR portions and dissonances by LPC also.

II. LITERATURE SURVEY

1) Speech Compressive Sampling Using Approximate Message Passing and a Markov Chain Prior.

Exercising compressive slice (CS), a stingy signal can be efficiently recovered from its far lower samples than that demanded by the Nyquist – Shannon slice theorem. To break this problem, we propose a system combining the approximate communication end (AMP) and Markov chain that exploits the dependence between a speech signal's modified separate cosines transfigure (MDCT) portions. A turbo frame, which alternately iterates AMP and belief propagation along the Markov chain, is employed to reconstruct the speech signal from CS samples. In addition, a constraint is set to the turbo replication to help the new system from divergence. Extensive trials show that, compared to other traditional CS styles, the new system achieves an advanced signal- to- noise rate and an advanced perceptual evaluation of speech quality (PESQ) score. The new system also achieves an analogous speech enhancement effect to the state- of- the- art system.

2) Speech Enhancement Using Deep Learning AVE-NET.

Utmost approaches to speech enhancement focus solely on audio features to produce adulterants or transfer functions that convert noisy speech signals to clean bones multitudinous speech- related approaches have integrated visual and audio data to achieve further effective speech- processing performance. Trials are conducted using the recently released NTCD- TIMIT dataset and the GRID corpus. The speech signal is heavily corrupted by noise. Our system integrates the capability of PCA toed-correlate the portions by rooting a direct relationship with sea packet analysis to decide point vectors used for speech improvement. This allows us to operate with an accessible loss function on these new portions, removing the noise without demeaning the speech. Also, the enhanced speech attained by the inverse sea packet transfigure is perished into three subspaces low rank, meagre, and the remainder noise factors. Eventually, we calculate the factors as an isolation problem. The performance evaluation shows that our system provides an advanced noise reduction and a lower signal deformation indeed in largely noisy conditions without introducing artefacts.

3) Speech Enhancement, Databases, Features, and Classifiers in Automatic Speech Recognition: A Review.

This paper gives a comprehensive introduction to speech production and perception in Human beings, discusses the challenges and, applications, framework of Automatic Speech Recognition (ASR) systems, and provides a detailed description of the architecture of ASR. The paper highlights the problem definition, general objectives, and scope of ASR research. A systematic way of designing a speech corpus for Indian languages is presented in this paper. The results convey that the proposed method performs better for both databases for various SNR conditions.

4) Model-Based Speech Enhancement for Time-Varying Noises.

Our work introduces a trainable speech improvement fashion that can explicitly incorporate information about speech signals' long-term, time- frequency characteristics before the improvement process. We compare noise spectral magnitude from available recordings from the functional terrain and clean speech from a clean database with fusions of Gaussian pdfs using the Anticipation-Maximization algorithm (EM). Latterly, we apply the Bayesian conclusion frame to the demoralized spectral portions, and by employing Minimum Mean Square Error Estimation (MMSE), we decide an unrestricted- form result for the spectral magnitude estimation task. We estimate our fashion with a focus on real, largely nonstationary noise types (passing- by aircraft noise) and demonstrate its effectiveness at low SNRs.

III. PROPOSED METHODOLOGY

The proposed methodology of speech enhancement based on approximate message passing (AMP) involves leveraging several algorithms, namely Independent Component Analysis (ICA), Principal Component Analysis (PCA), Fast-ICA, and Kalman filtering. AMP, an iterative signal processing technique for sparse signal recovery, is at the core of this approach. In the context of speech enhancement, ICA plays a vital role in separating the mixed signals into their independent source components, enabling the identification and attenuation of noise in the observed noisy speech signal. PCA aids in dimensionality reduction, helping to identify dominant speech-related components and discard noise-related ones. Fast-ICA, a quicker implementation of ICA, efficiently extracts the independent components from the mixed signals. Lastly, Kalman filtering is employed to model speech dynamics over time, resulting in improved noise reduction by updating the estimates of speech and noise components. The combination of these techniques within the AMP framework holds promise for enhancing speech quality and reducing noise interference in practical applications. It's important to note that this field is dynamic

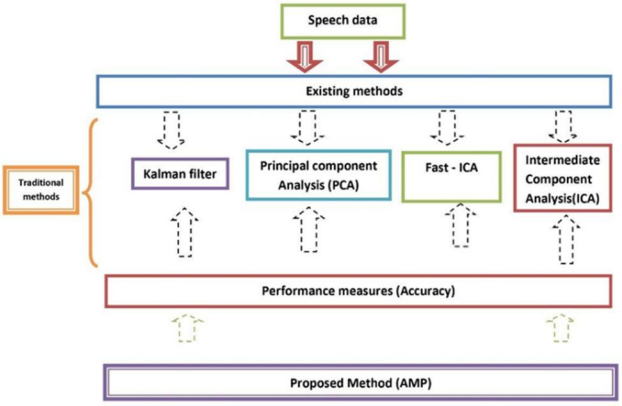


Fig.1: Architecture of the proposed system

A. *Dataset*

S.no	Audio	File path	Start segment	End segment	hertz	decibles	Sample Frequency
1	source1.wav	Click here!!	0	0.5	1895.36 Hz	0.35 dB	8000 Hz
2	source2.wav	Click here!!	0.1	0.3	134.56 Hz	0.49 dB	8000 Hz
3	source3.wav	Click here!!	0.1	0.6	34.88 Hz	0.07 dB	8000 Hz
4	source4.wav	Click here!!	0	0.4	166.88 Hz	2.41 dB	8000 Hz
5	source5.wav	Click here!!	0	0.4	97.44 Hz	0.63 dB	8000 Hz

Table I: Speech sound dataset

1) The proposed methodology involves using the Approximate Message Passing (AMP) framework for speech enhancement.
2) The algorithms employed in this approach are Independent Component Analysis (ICA), Principal Component Analysis (PCA), Fast-ICA, and Kalman filtering.
3) The "centerRows" function is used to compute the sample mean along each feature (column) of the input data matrix and then center the data by subtracting the mean from each sample.

4) The "Demo ICA" demonstrates the application of ICA algorithms on audio signals, separating independent components from mixed observations using Fast-ICA with different contrast functions.
5) "Demo PCA1" showcases Principal Component Analysis on Gaussian data, transforming high-dimensional data into a lower-dimensional space while preserving important patterns and variances.
6) "Demo PCA2" illustrates the effectiveness of PCA in decomposing correlated multivariate Gaussian samples into orthogonal, maximal variance components.
7) "Evaluation Metrics" module provides quantitative assessment of the performance of signal separation algorithms, aiding the comparison of accuracy and quality against ground truth data.
8) "Fast ICA" implements Independent Component Analysis using the Fast-ICA algorithm, extracting independent components from the given data by maximizing non-Gaussianity measures.
9) "KIca" uses the max-kurtosis ICA algorithm to separate independent components from observed fusions.
10) "Load Audio" simplifies loading and preparing audio data for analysis in MATLAB.
11) "Normalize Audio" scales audio data to have a unit norm, useful for consistent and meaningful audio signal processing.

IV. IMPLEMENTATION

1) *VAD Algorithm:* Detects speech in audio by analyzing energy and spectral features, minimizing false positives/negatives, and enhancing speech-related systems.
2) *Kalman Filtering:* Recursive algorithm for estimating dynamic system states from noisy measurements, widely used in navigation, robotics, and speech enhancement.
3) *FastICA:* Efficient algorithm for Independent Component Analysis, effectively separating independent sources from mixtures in audio and image processing.
4) *ICA Algorithm:* Statistically separates mixed signals into independent sources, used in blind source separation, noise reduction, and neuroscience.
5) *PCA Algorithm:* Reduces data dimensionality while preserving essential patterns, commonly applied in data visualization, compression, and feature extraction.
6) *Applications:* VAD optimizes speech recognition and compression, Kalman filtering handles real-time system estimation, FastICA and ICA aid in source separation, and PCA simplifies data representation and analysis.

V. EXPERIMENTAL RESULTS

A digital audio expander is an audio processing technique employed to enhance the dynamic range of an audio signal. By amplifying low-level signals while preserving the integrity of louder segments, it brings out subtle details and improves audibility in quieter portions of the audio, ultimately resulting in enhanced sound characteristics. Additionally, it effectively reduces the perception of background noise, leading to a clearer and more immersive listening experience. When applied to noisy speech signals, digitally expanded noisy speech selectively amplifies low-level speech components, making the speech more discernible and significantly improving intelligibility, particularly in noisy environments. The process involves carefully boosting specific speech features while minimizing the amplification of noise, ensuring that the speech remains clear and intelligible to listeners.

In speech analysis, Autoregressive (AR) coefficients are utilized in autoregressive models to predict a data point in a time series based on its past values. These coefficients represent the weights assigned to previous observations, defining the linear combination that characterizes the model. By employing AR coefficients, one can effectively model the underlying patterns and dynamics of a speech signal, enabling precise prediction and analysis. A filtered speech frame, on the other hand, refers to a segment of speech that has undergone a filtering process to remove unwanted components, such as noise or distortion. This filtering improves speech quality and intelligibility by enhancing specific characteristics and removing undesirable elements that could hinder effective communication. Various filtering techniques, such as adaptive filters, spectral subtraction, or digital audio effects, can be applied to achieve the desired enhancements. Pitch power, also known as pitch strength, plays a crucial role in speech analysis by quantifying the perceived prominence or strength of the fundamental frequency in a speech signal.

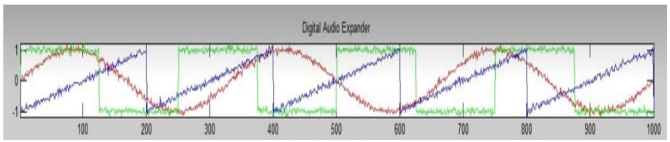


Fig. 2: Digital Audio Expander

©IJRASET: All Rights are Reserved

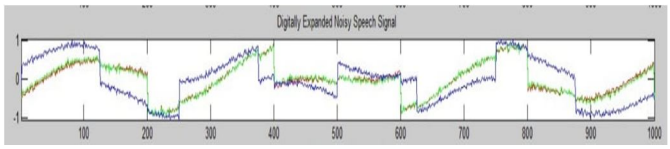


Fig. 3: Digitally expanded noisy speech signal

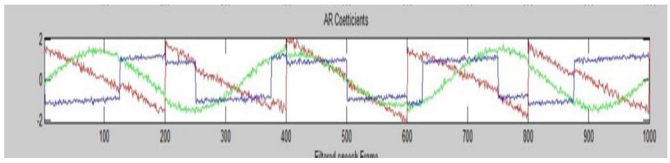


Fig. 4: AR coefficients

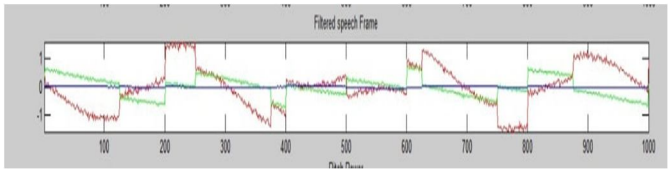


Fig. 5: Filtered Speech Frame

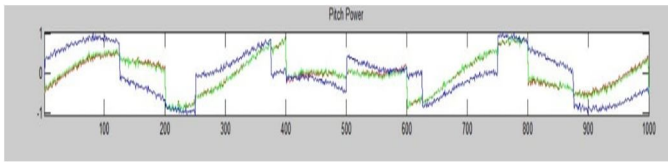
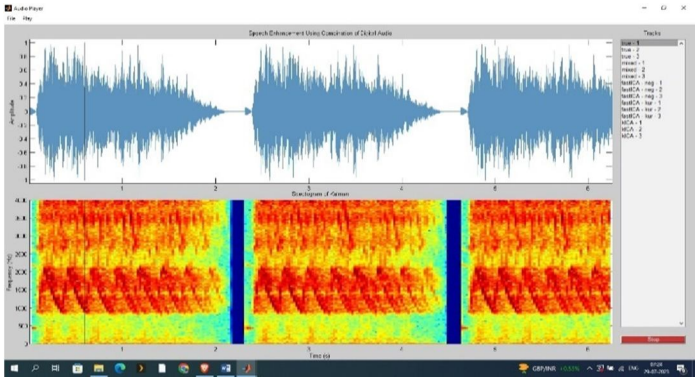
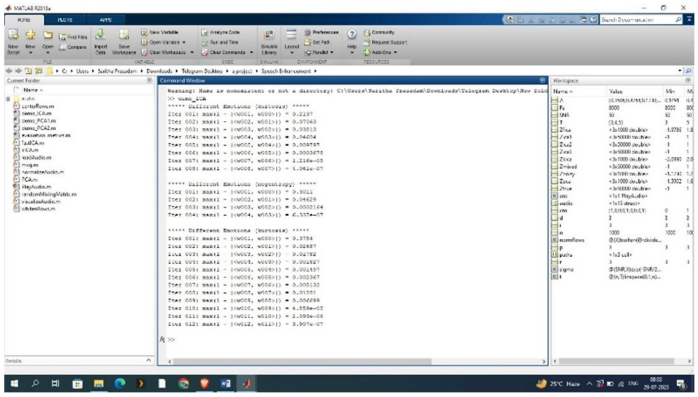


Fig. 6: Pitch power


Fig. 7: spectrogram of the speech sound by the audios


Fig. 8: Utmost Iterations are made for filtering the noise

©IJRASET: All Rights are Reserved

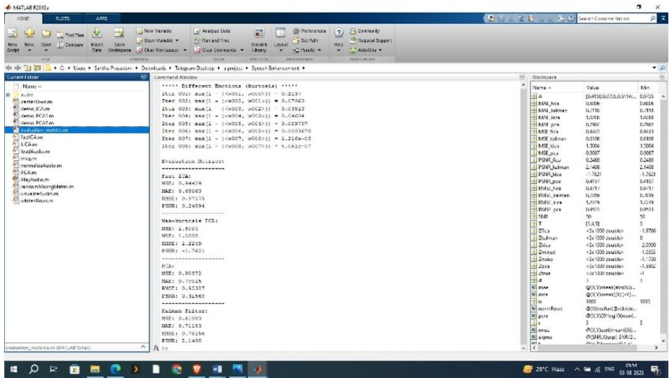


Fig. 9: evaluation metrics of the taken audios

Algorithms	Kalman filter	Fast ICA	ICA	PCA
M.S.E	0.61	0.94	1.50	0.90
M.A.E	0.71	0.68	1.02	0.79
R.M.S.E	0.78	0.97	1.22	0.95
P.S.N.R	2.14 dB	0.24 dB	-1.76 dB	0.41 dB

Table II: evaluation metrics of the algorithms

VI. CONCLUSION

The proposed method addresses the drawbacks of basic methods like spectral subtraction and Wiener filter, as well as the complexity and time-consuming nature of other Kalman filter-based approaches.

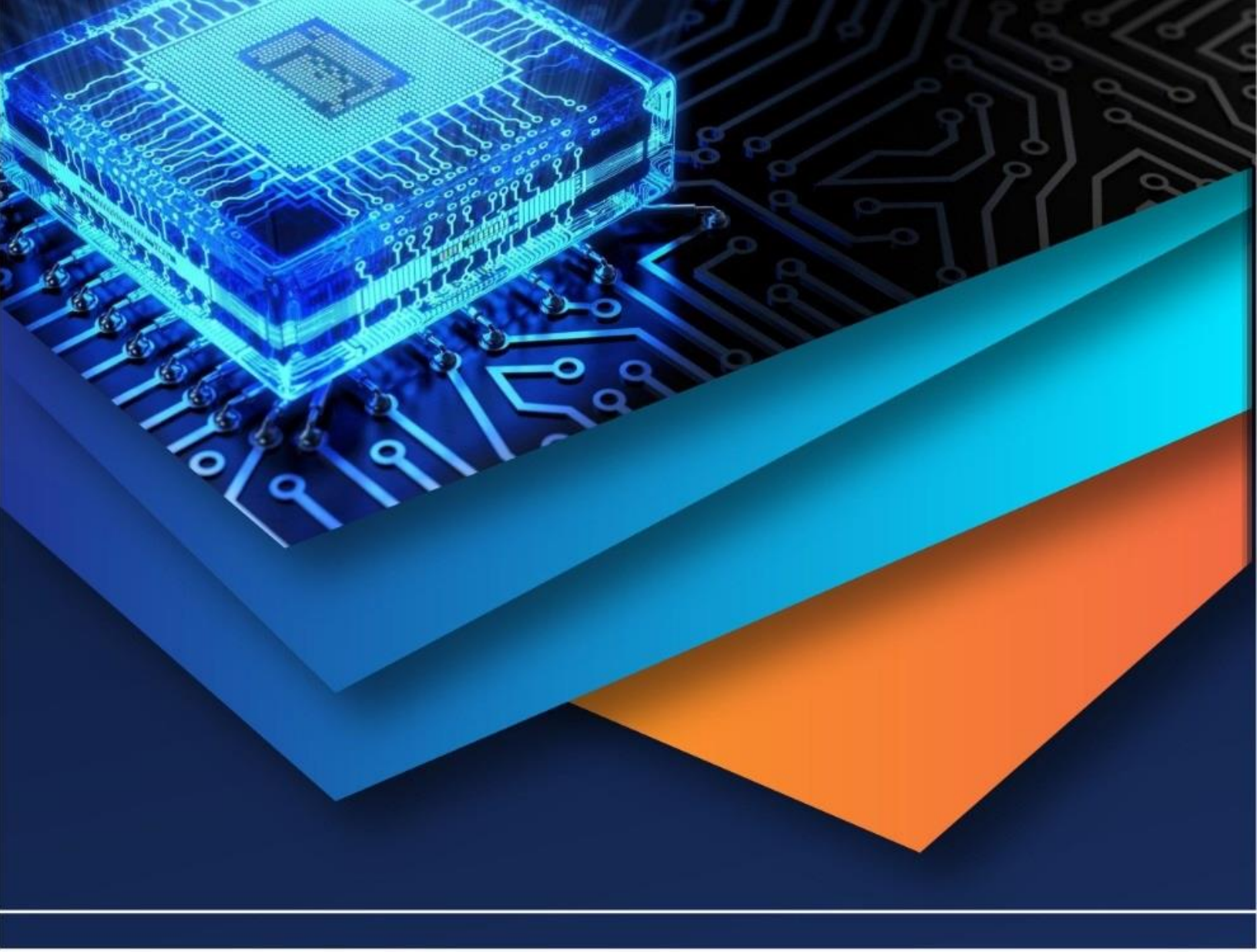
Through simulations in MATLAB and comparisons of input-output SNR values, the performance of the proposed method is evaluated. The results show that the proposed method outperforms conventional methods, including ICA, Fast ICA, PCA, and Kalman Filter, in terms of output SNR values for both stationary and non-stationary signals. The combination of Digital Audio Effecting and the improved Adaptive Kalman Filter allows for effective noise reduction while considering the perceptual aspects of the human ear, leading to superior speech quality and intelligibility.

While the study focuses on ICA, Fast ICA, PCA, and Kalman Filter algorithms as points of comparison, the proposed method's unique combination of Digital Audio Effecting and improved Adaptive Kalman Filter sets it apart from these individual techniques. The innovative integration of these components highlights the paper's contribution to advancing speech enhancement research and offers potential for future investigations into novel hybrid approaches.

In conclusion, the proposed speech enhancement method based on approximate message passing, incorporating the ICA, Fast ICA, PCA, and Kalman Filter algorithms, offers a promising approach to significantly improve speech quality in noisy environments. The method's ability to effectively reduce background noise while minimizing distortion and its superior performance compared to traditional techniques make it a valuable contribution to the field of speech enhancement.

REFERENCES

[1] E. Dong,X. Pu, " Speech denoising predicated on perceptual weighting sludge," 9th International Conference on Signal Processing, Leipzig, Germany, 2008, pp 705- 708.
[2] K.Sakhnov, E.Verteletskaya, B. Šimák, "Dynamical Energy- Grounded Speech Silence Sensor for Speech Enhancement Applications." In World Congress of Engineering 2009 Proceedings. Hong Kong, 2009, pp. 801- 806.
[3] "Speech Enhancement" byJ. Benesty,J. Chen,Eds., andS. Makino, Springer, Berlin, 2005.
[4] S.Boll," suppression of audial noise in speech using spectral deduction," IEEE Transaction on Speech, Signal Process. Volume.ASSP- 27, no. 2pp.113- 120, 1979
[5] T.V.SreenivasandP.kirnapure," Codebook constrained Wiener filtering for speech enhancement," IEEE Trans. Speech and Audio Processing, sep1996.
[6] "Digital Audio Signal Processing," second edition, by Udo Zolzer.
[7] Hu.Y and Loizou.P,"Evaluation of objective quality measures for speech enhancement," IEEE Trans. On speech and audio processing, 2008.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)