SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V10I1P113

Volume 10 Issue 1, 138-147, January 2023

Original Article

Discrete Wavelet Transform with Thresholding: An Effective Speech De-Noising Algorithm

Jagadish S.Jakati[1], Veereshkumar Mathad[2], Kiran S Nandi[3], Anandraddi Naduvinmani[4]

[1,3,4]*Department of E & C Engineering, S G Balekundri Institute of Technology, Shivbasvnagr, Belagavi-10, VTU, Karanataka, India*
[2]*Department of Electrical EngineeringMaharaj College of Engg, Mahagaon, Maharashtra, India.*

[1]*Corresponding Author : jagadishjs30@gmail.com*

Received: 03 December 2022 Revised: 10 January 2023 Accepted: 19 January 2023 Published: 29 January 2023

Abstract - *The research of speech augmentation has become increasingly popular in the domain of speech processing. It mainly concentrates on removing the voice stream's additive background noise, significantly degrading speech interpretability. The objective of speech enhancement is to eliminate additive noise from the speech signal and restore the original signal. There are presented methods for improving speech based on speech and noise signals' perceptual, auditory, or statistical limitations. Predicting the features of the voice signal and any background noise is quite difficult in a decent environment. Speech processing is challenging due to the absence of a specific framework for the speech signal and a cognitively important distortion scale. Speech transmissions are also, by nature, non-stationary. Consequently, adaptive estimate methods that don't require an explicit predictive method for the underlying signal statistics typically overlook changes. Therefore, by utilizing voice enhancement techniques, signal noise can be somewhat decreased. Additionally, there is a trade-off between the amount of noise suppressed and the irregularities in the voice signal produced. This study aims to provide an efficient method for examining voice augmentation techniques. Another problem is the simplicity with which noise-suppression algorithms can be applied in mobile phones and digital hearing aids. New strategies are needed to improve the effectiveness of speech enhancement technologies in light of the aforementioned limitations. Due to their excellent efficiency, transform domain filters are frequently used in this study's speech improvement process.*

Keywords - *Speech processing, DCT, Discrete Wavelet Transform, Thresholding, Signal to noise ratio, MoS. LLR ISD, etc.*

1. Introduction

Speech is the vocalised form of human communication. Its foundation is the syntactic integration of terms from extensive dictionaries. Enhancement is the increase or improvement of quality, value, or extent. The aim of improving speech is to significantly enhance the quality and comprehensibility of a compromised spoken signal using audio signal processing methods. A variety of noises impedes speech. Speech intelligibility and quality suffer as a result. The most important aspect of enhancing speech is recovering speech that has been inhibited by noise. Voice augmentation is used in a number of applications, including teleconferencing, VoIP, mobile phones, voice recognition, and hearing impairment. A speech augmentation system needs to provide excellent service for all voice signals. The user chooses sample functions randomly as part of the input method for the voice augmentation system. By definition, noise is stochastic. In light of this, the statistical estimation problem that depends on a single random process and noise is the speech enhancement problem. The mathematical theories of signal and noise, as well as a distortions metric that assesses the similarity between the clean signal and its anticipated version, are necessary for estimation theory. These two essential elements of estimate theory are not readily available for voice signals. A lack of a framework for reliable speech signals and a distortion metric with cognitive significance causes the problems. Additionally, the speech signals are not continuous at all times. The result is Adaptive estimate methods that don't require a particular forecast method for the signal frequently miss changes in the fundamental statistics of the signal.

2. Literature Survey

This gives a brief description of the methods currently in use to denoise speech signals. The coherence function is estimated without the input signals relying on prior noise statistics in this paper's [3] new method for multi-channel speech enhancement. A gain function is used by the discrete wavelet transform (DWT). When used in a home with interfering speakers, this strategy produced good results when numerous noise kinds corrupts speech.

The use of wavelet de-noising on the speech input of the MFCC feature extraction method was presented in this research [4]. The MFCC displays on noisy signals are improved using the wavelet-based de-noising method. They employed 120 speech samples, of which 90 served as testing samples, and 30 served as references. These wavelet-based techniques can improve the system's speech recognition accuracy for signals with SNRs of 0–10 dB.

In this study [5], a DWT-based technique has been employed to eliminate the noise from the audio stream. De-noising employs both hard and soft thresholding. This technique can be utilised for real-time processing and achieve effective results.

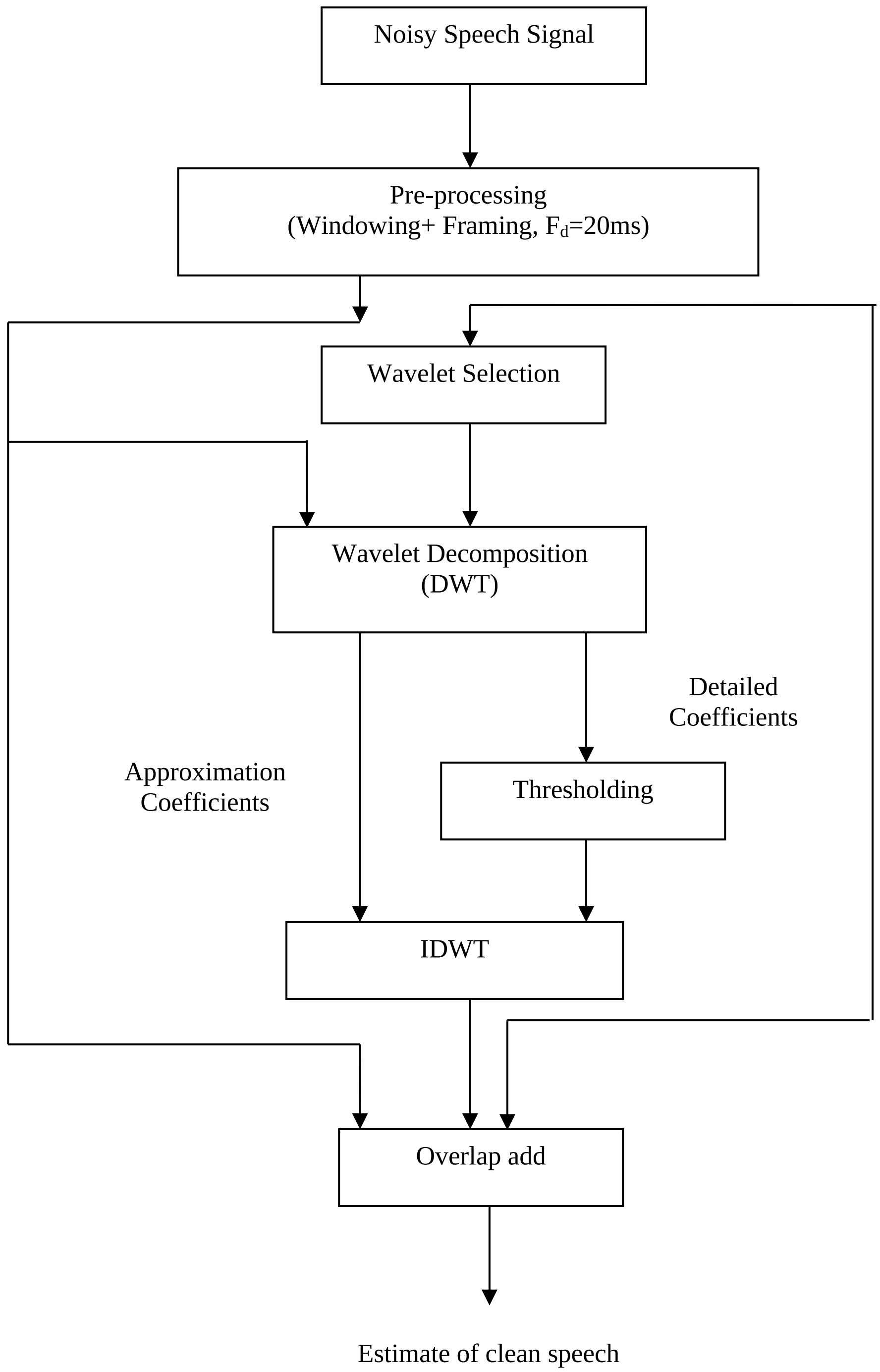


Fig. 1 Proposed algorithm of speech de-noising employing discrete wavelet transform and thresholding

In this study [6], DWT is employed in place of STFT to analyse the E-DATE algorithm's noise power spectrum. The creative approach successfully recovered the STFT of the input speech signal by wavelet decomposing the framed signal and thresholding the precise coefficients. This approach offered the best outcomes in terms of objective evaluation measures like the signal-to-noise ration and MOS scores.

They proposed speech enhancement using the DWT and discrete cosine transformation in this study [7], followed by speech de-noising using a packet wavelet. Using MSE and PSNR, determine the acoustic compression's quality. Wavelet transforms coupled with DCT yield amazing results. The DCT approach has received good results estimates from DWT, with a PSNR of roughly 50 dB.

They proposed the integrated wavelet threshold and MMSE-LSA, this study's improved speech de-noising technique for low SNR speech signals [10]. The proposed method significantly outperformed voice intelligibility simulations than wavelet threshold de-noising and EMD-based wavelet threshold de-noising methods. Speech signal de-noising is accomplished using a discrete wavelet packet transform technique in this research [11]. Samples of noisy speech signals contaminated by white Gaussian noise from 0dB to +15dB are denoised using both hard and soft thresholding. They concluded that soft thresholding is more effective for output SNR values than hard thresholding.

A method was employed for speech de-noising based on the based on principle component review suggested in this study [12]. This principal component analysis-based de-noising technique had a higher impact. Additionally, the principle component denoising technique's signal waveform was more complete and similar to the original speech signal.

3. Proposed Method

The suggested solution for effective speech de-noising employing discrete wavelet transform and thresholding is presented in this part.

Step 1: Initially, noisy speech signal from the source is pre-processed by the window and overlapping method.

Step 2: Windowing is done by the hamming window; the size of the window can be taken as10-30msec, and the frame size is 20msec.as the disadvantage of the hamming window, there may be the possibility of spectral leakage at the edge of the window so to avoid this 50% overlapping is done.

Step 3: Wavelet selection process involves the methods for choosing the best wavelet for speech signal. Here we select Coiflet1.

Step 4: After choosing a wavelet, wavelet decomposition is performed using a discrete wavelet transform that splits the wavelet into many sets, each set comprising a time series of coefficients that indicate the signal's temporal history in the appropriate frequency band. (i.e. approximation and detailed coefficients).

Step 5: The approximation coefficients directly compute SNR using inverse discrete wavelet Transform, and detailed coefficients apply the suitable thresholding technique and compute SNR using discrete inverse wavelet Transform.

Step 6: After computing SNR, all samples can recombine into overlapping chunks called frames. The frames are then reassembled to estimate a clean speech signal. Each frame is 50% overlapped.

4. Performance Parameters

The parameters that are used to determine the performance [29] of the suggested method are covered in this section.

4.1. Increment in Segmental SNR

As a result of the speech-enhanced signal's non-stationarity, its energy fluctuates at random. As a result, evaluating voice quality while treating the entire signal as one may not be correct. Hence Segmental SNR is created by individually calculating and combining the SNR of each frame segment. The time or transform domains can be used to run this test. The relationship can be used to calculate segmental SNR.

$$SNR_{Seg} = \frac{10}{1}\sum_{i=0}^{I-1} log_{10} \frac{\sum_{n=N_i}^{N_i+N-1} x^2(n)}{\sum_{n=N_i}^{N_i+N-1}(x(n)-\hat{x}(n)^2} \quad (1)$$

Where, N: Frame length.
L: Number of frames.
x(n): Original noisy speech.
x^(n):- Processed speech signal.

After obtaining the segmental SNR value, we must compute the increase in segmental SNR, also known as the SNR of the enhanced speech signal from the noisy speech signal. The main drawback of this evaluation approach is the potential for receiving negative values during the silence intervals.

4.2. Log Likelihood Ratio (LLR)

Based on linear predictive coding, the test's objective measurements are employed. The tilt in phase between the unprocessed speech signal and the enhanced speech signal is calculated using the LLR test. This value represents the distortion that was added throughout the processing phase. LLR frequently has a value of 2. The following is the LLR

Calculation formula

$$D_{LLR}(a_e, a_c) = log_{10} \frac{a_e R_c a_e{}^T}{a_c R_c a_e{}^T} \quad (2)$$

Where, ac: LPC vector of the clean speech signal.
a_e: LPC vector of the enhanced speech signal.
R_c: Autocorrelation matrix of the clean speech signal.

4.3. Itakura–Saito Spectral Distance (ISD)

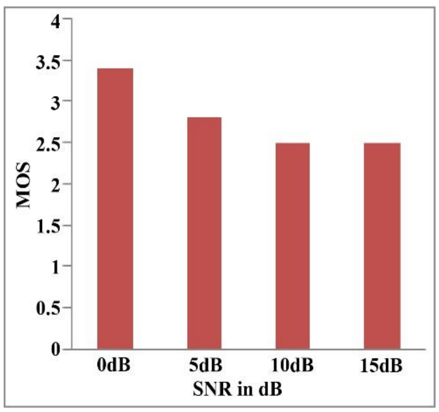
Another PLC-based test of evaluation using Objectives measurements is the Itakura-Saito spectral distance. The spectral envelope of the augmented speech signal and the enhanced speech signal differs as a result. Its value is typically less than 100. The following is the equation to compute ISD:

$$D_{ISD}(a_e, a_c) = \frac{G_c}{G_e}\frac{a_e R_c a_e^T}{a_c R_c a_e^T} + log_{10}\frac{G_c}{G_e} - 1 \qquad (3)$$

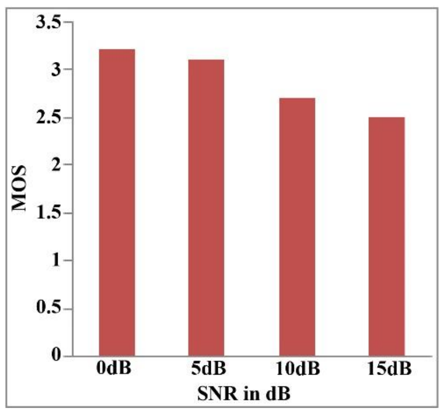
Where Ge and Gc stand for processed speech's and clean speech's respective linear predictive coding improvements.

5. Results and Discussions

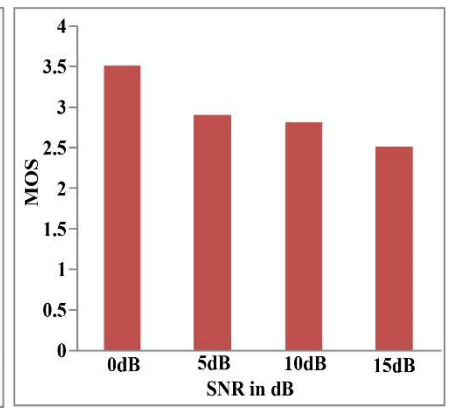
5.1. Subjective Listening Test



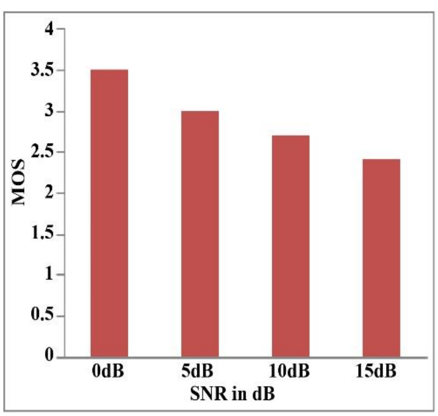
Babble noise



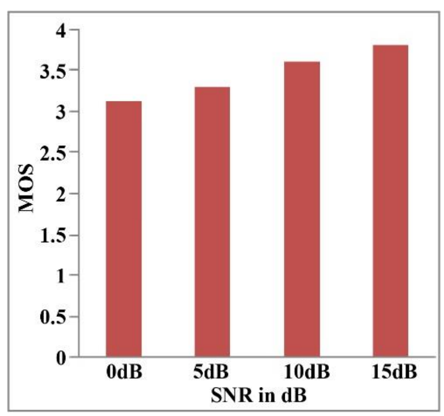
Airport noise



Station noise



Restaurant noise



Additive white Gaussian noise

Fig. 2 Bar graphs of MOS of normal hearing subjects for DWT and Thresholding method, enhancement method for different noise sources (Babble, Airport, Station noise, Restaurant, AWGN).

Observations of the subjective listening test

- From the bar graphs, we can observe that there is a considerable upsurge in the MOS of the proposed algorithm.
- The mean opinion score value obtained from normal hearing listeners is more than that obtained by the subjects with hearing impaired problems.
- From the above graphs, we can also observe that the proposed algorithm performs well at 0dB for all kinds of noise signals but consistently works well at all SNR values when the noise signal is AWGN.

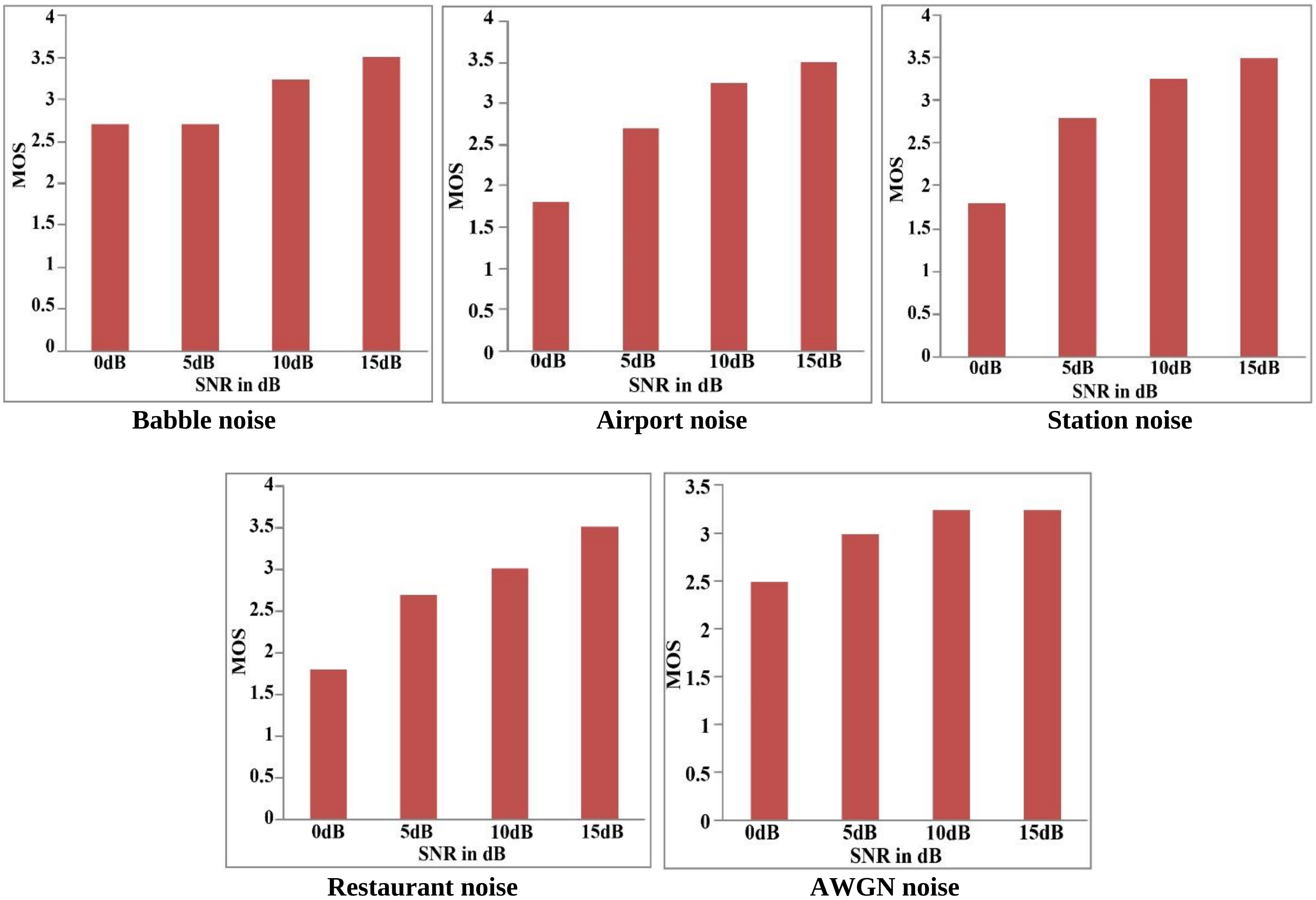
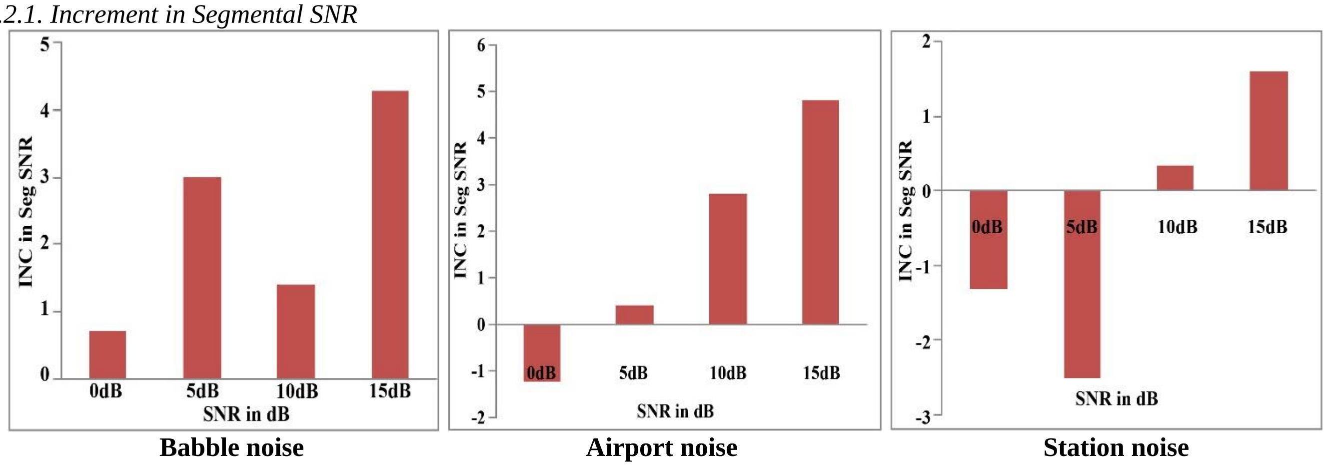


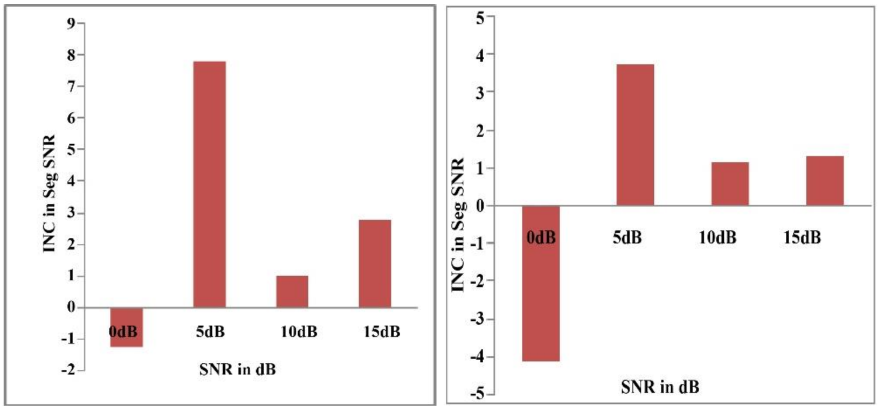
Fig. 3 Bar graphs of MOS of subjects with impaired hearing for DWT and Thresholding method, enhancement method for different noise sources (Babble, Airport, Station noise, Restaurant, AWGN)

5.2. Objective Measures

Three tests are conducted in the objective measures evaluation method, Increment in segmental SNR, Log-likelihood ratio (LLR) and Itakura-Saito distance. By using mathematical equations, values of all three methods are plotted for the Multilevel Discrete Wavelet Transform method once again for different kinds of noise signals.

5.2.1. Increment in Segmental SNR





Restaurant noise **AWGN noise**

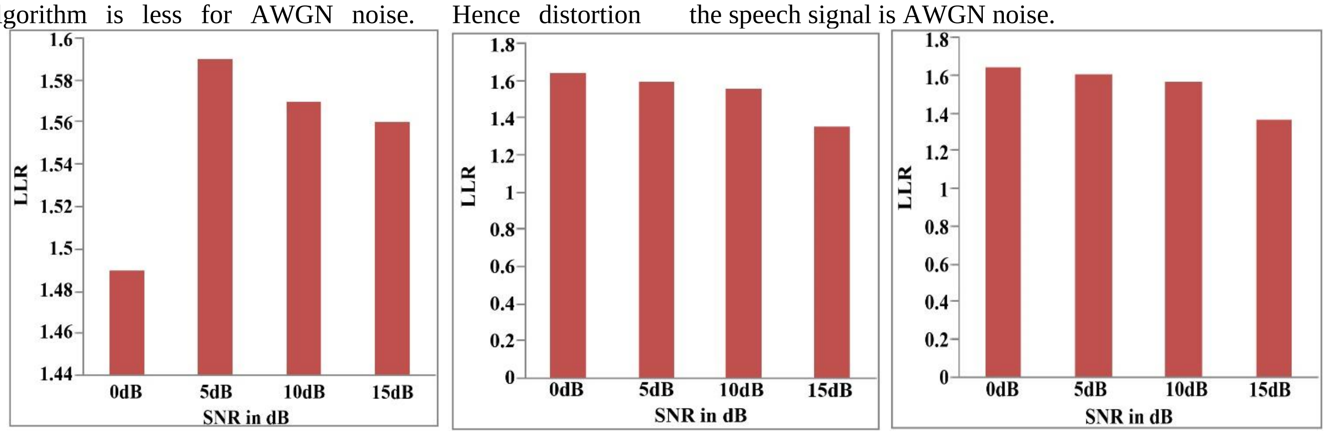
Fig. 4 Bar graphs of increment in segmental SNR for three Enhancement method Discrete Wavelet Transform and Thresholding method, for five different noise sources (Babble, Airport, Station, Restaurant noise, AWGN noise).

Observations of the test conducted

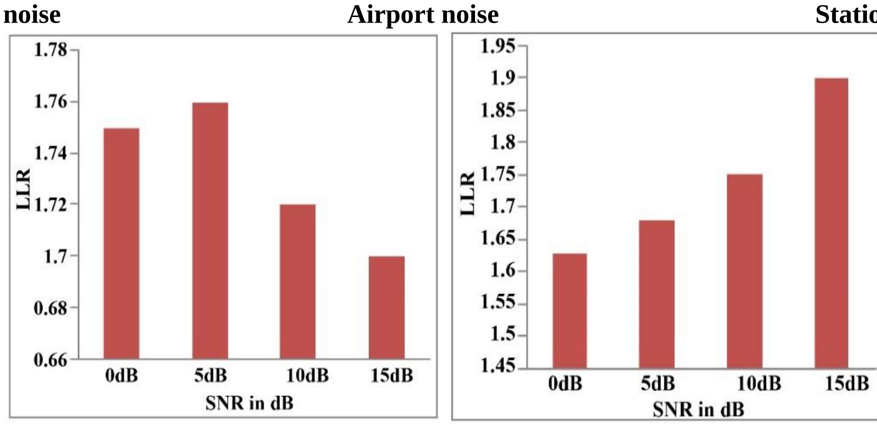
- From the bar graphs plotted above, we can observe that increment in segmental SNR value has been increased in the proposed algorithm in speech enhancement methods.
- As the SNR of a noisy speech signal rises, the increment in segmental SNR decreases because such noisy speech signals include less noise, but more noise is unnecessarily eliminated owing to processing, which lowers the performance of the suggested approaches.
- From the 5.2.1 graphs, we can observe that there is an increment in SNR value from 0dB to 15dB for babble noise and AWGN noise. Hence a proposed method performs well when noise present in the speech signal is babble and AWGN noise.

5.2.2. Log – Likelihood Ratio

It has been noted that the LLR of The appropriate algorithm is less for AWGN noise. Hence distortion introduces by the process is minimum when noise appears in the speech signal is AWGN noise.



Babble noise **Airport noise** **Station noise**



Restaurant noise **AWGN noise**

Fig. 5 Bar graphs of LLR for DWT and Thresholding method enhancement processes for different kinds of noise sources(Babble, Airport, Station, Restaurant noise, AWGN noise).

5.2.3. Itakura – Saito Distance

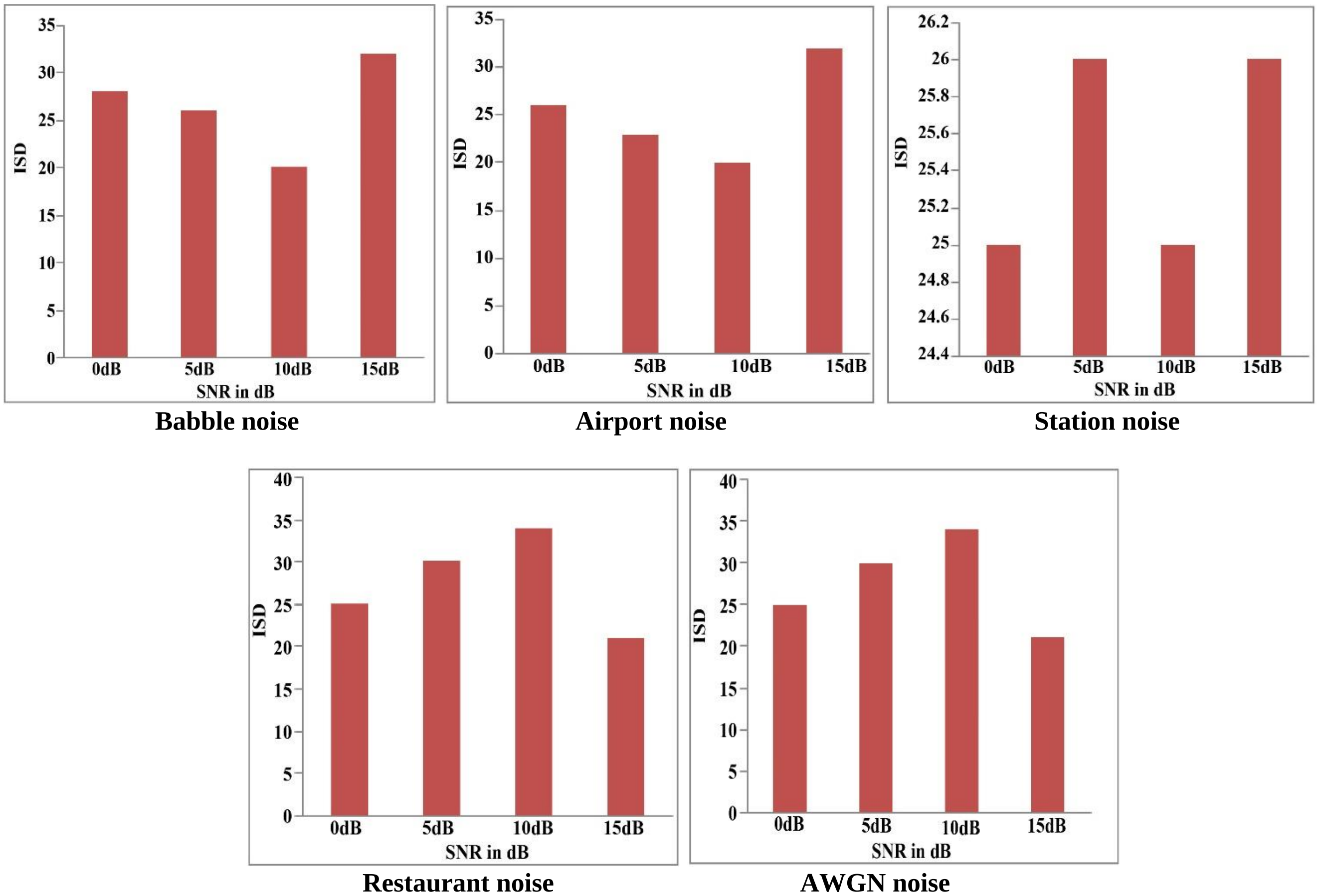
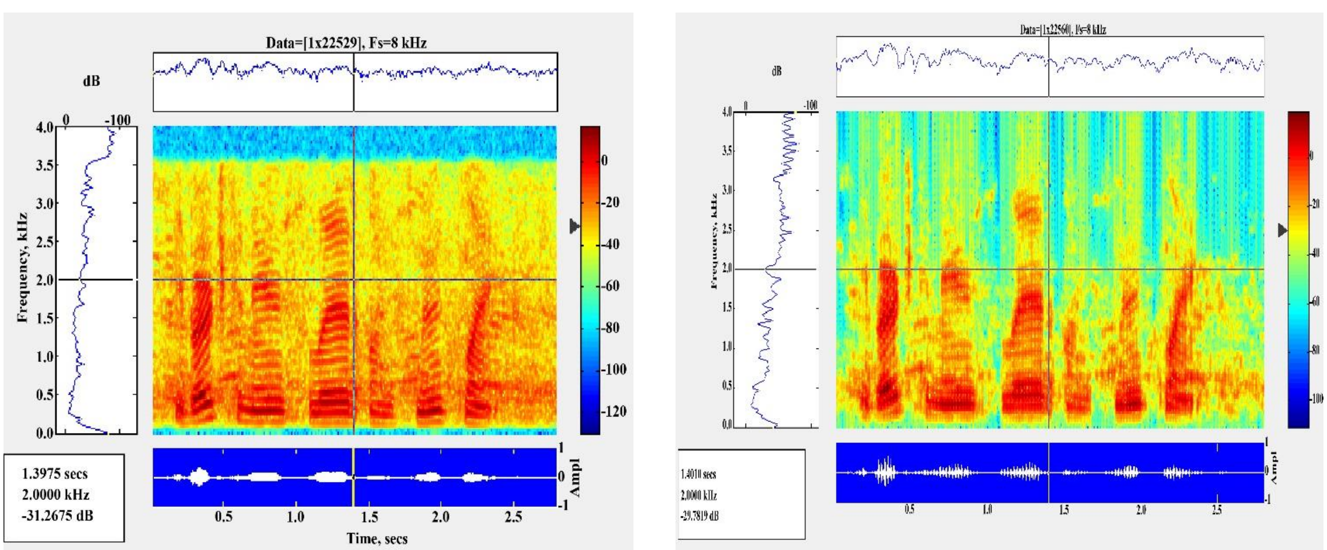
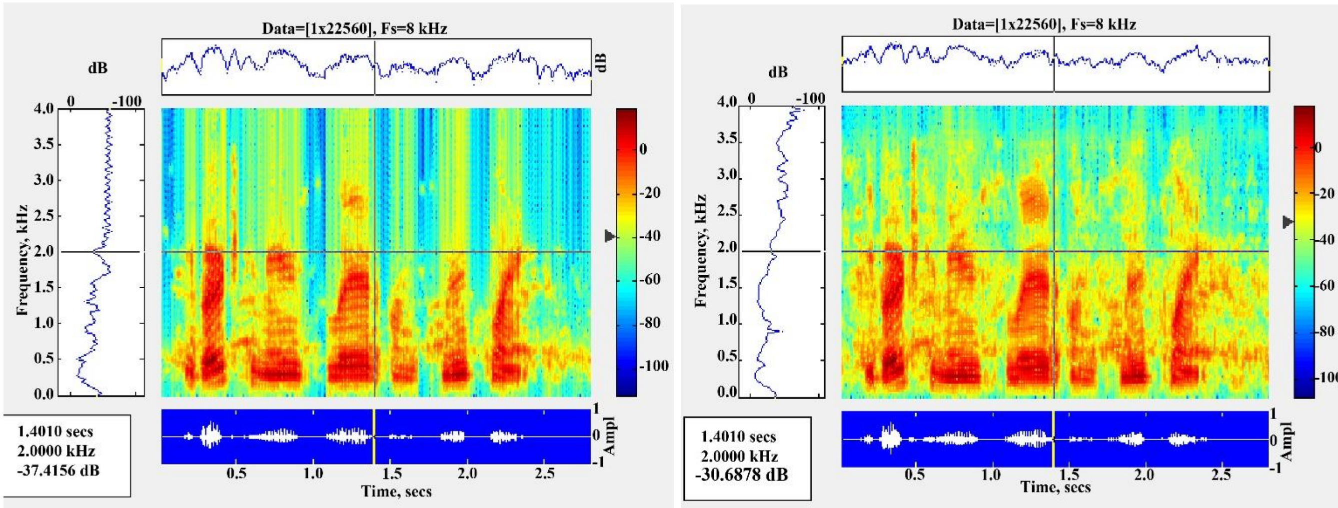


Fig. 6 Bar graphs representing Itakura-Saito distance for DWT and Thresholding method for different kinds of noise signals (Babble, Airport, Station, Restaurant, AWGN noise).



(a) **(b)**



(c) (d)

Fig. 7 Spectrogram images of the noisy and enhanced speech signals (a) 15dB noisy speech signal despoiled by babble noise, (b) speech signal enhanced by Discrete Wavelet Transform method Using thresholding, (c) speech signal enhanced by Discrete Wavelet Transform method Using thresholding, (d) speech signal enhanced by the Discrete Wavelet Transform method Using thresholding.

From spectrogram analysis, we can state that the enactment of the projected work is superior to the existing speech enhancement methods, especially when noise signals are babble and AWGN noise signals.

6. Conclusion and Future Scope

The foundations of the efficient speech decoding technique suggested in this article are thresholding and the discrete transformation of wavelets. The full voice signal is divided into smaller frames for proper processing, which are then subjected to various DWT-IDWT algorithms to remove noise components. Through the use of inverse DWT, thresholding lowers the level of noise utilised to denoise the data. The proposed technique may denoise speech more effectively than existing ones, thanks to the usage of DWT-IDWT and thresholding, as shown by the performance metrics. Higher-order filter banks with dynamic taping factor features will be considered in the future for more effective de-noising.

Conflicts of Interest

A research topic is published in a journal, an authors' conflict of interest disclosure statement stating that they have none to disclose must also be included. Each work submitted for publication in the journal must be accompanied by a proclamation from the scholars stating they have no conflicts of interest to mention. Unsigned editorials need never be published by a journal since someone with a conflict of interest apparently writes them.

Acknowledgments

The authors express their gratitude to Dr. B. R. Patagundi, Principal of the S. G. Balekundri Institute of Technology in Belagavi, and Dr. S. H. Sawant, Principal of the Sant Gajanan Maharaj College of Engineering in Mahagaon, Maharashtra, India. SGBIT Belagavi and SGMC Mahagaon, Maharashtra Belagavi employees, thank you for your kind assistance with this research.

References

[1] Tusar Kanti Dash, and Sandeep Singh Solanki, “Comparative Study of Speech Enhancement Algorithms and Their Effect on Speech Intelligibility,” *2017 2nd International Conference on Communication and Electronics Systems, IEEE,* pp. 270-276, 2017. *Crossref,* http://doi.org/10.1109/CESYS.2017.8321280

[2] Anamika Baradiya, and Vinay Jain, "Speech and Speaker Recognition Technology using MFCC and SVM," *SSRG International Journal of Electronics and Communication Engineering,* vol. 2, no. 5, pp. 6-9, 2015. *Crossref,* https://doi.org/10.14445/23488549/IJECE-V2I5P105

[3] Kristian Timm Andersen, and Marc Moonen, “Robust Speech-Distortion Weighted Interframe Wiener Filters for Single-Channel Noise Reduction,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* vol. 26, no. 1, pp. 97-107, 2018. *Crossref,* http://doi.org/10.1109/TASLP.2017.2761699

[4] Gerald Enzner, and Philipp Thüne, “Robust MMSE Filtering for Single-Microphone Speech Enhancement,” *2017 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE,* pp. 4009-4013, 2017. *Crossref,* http://doi.org/10.1109/ICASSP.2017.7952909

[5] Yasser Ghanbari, and Mohammad Reza Karami-Mollaei, “A New Approach for Speech Enhancement Based on the Adaptive Thresholding of the Wavelet Packets,” *Speech Communication*, vol. 48, no. 8, pp. 927-940, 2006. *Crossref*, https://doi.org/10.1016/j.specom.2005.12.002

[6] Sumer Singh Singhwal, “Noise Reduction from Speech Signal Using MATLAB and Wavelet Transform,” *The Engineering Journal of Application & Scopes*, vol. 1, no. 1, 2016.

[7] Meet H. Soni, and Hemant A. Patil, “Novel Subband Autoencoder Features for Non-Intrusive Quality Assessment of Noise Suppressed Speech,” *Interspeech*, pp. 3708-3712, 2016. *Crossref*, https://doi.org/10.21437/Interspeech.2016-693

[8] Rajesh Kumar Dubey, and Arun Kumar, “Non-Intrusive Speech Quality Assessment Using Multi-Resolution Auditory Model Features for Degraded Narrowband Speech,” *IET Signal Processing*, vol. 9, no. 9, pp. 638-646, 2015. *Crossref*, https://doi.org/10.1049/iet-spr.2014.0214

[9] Dr. S.D. Apte, “Speech Enhancement in Hearing Aids Using Conjugate Symmetry Property of Short Time Fourier Transform,” *International Journal of Recent Trends in Engineering*, vol. 2, no. 5, pp. 346-351, 2009.

[10] S D Apte, and Shridhar, “Speech Enhancement in Hearing Aids Using Conjugate Symmetry of DFT and SNR-Perception Models,” *International Journal of Computer Applications*, vol. 1, no. 21, pp. 44-51, 2010. *Crossref*, https://doi.org/10.5120/58-650

[11] A. M. Mutawa, “Improving Patient Voice Intelligibility by using a Euclidian Distance-based Approach to Improve Voice Assistant Accuracy,” *International Journal of Circuits, Systems and Signal Processing*, pp. 329-339, vol. 14, 2020. *Crossref*, https://doi.org/10.46300/9106.2020.14.45

[12] Ch. D. Umasankar, and M Satya Sairam, “Performance Analysis of LMS, NLMS Adaptive Algorithms for Speech Enhancement in Noisy Environment,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 4, pp. 2330-2333, 2020. *Crossref*, https://doi.org/10.35940/ijitee.D1864.029420

[13] V.A.Mane et al., “Comparison of LDM and LMS for an Application of a Speech,” *International Journal on Signal Processing*, vol. 5, no. 4, pp. 130-141, 2011.

[14] M. A. Ali, and P. M. Shemi, “An Improved Method of Audio Denoising Based on Wavelet Transform,” *IEEE International Conference on Power, Instrumentation, Control and Computing*, pp. 1-6, 2015. *Crossref*, https://doi.org/10.1109/PICC.2015.7455802

[15] Pinki Sahil Gupta, “Speech Enhancement using Spectral Subtraction Type Algoritms: A Survey on Comparison,” *International Journal of Engineering and Computer Science*, vol. 4, no. 10, 2015.

[16] Prashanth Kannadaguli, and Vidya Bhat, "Phoneme Modeling for Speech Recognition in Kannada using Multivariate Bayesian Classifier," *SSRG International Journal of Electronics and Communication Engineering*, vol. 1, no. 9, pp. 1-4, 2014. *Crossref*, https://doi.org/10.14445/23488549/IJECE-V1I9P101

[17] J Indra et al., “A Modified Tunable–Q Wavelet Transform Approach for Tamil Speech Enhancement,” *IETE Journal of Research*, pp. 1-14, 2020.

[18] K.Sureshkumar, and Dr.P.Thatchinamoorthy, "Speech and Spectral Landscapes using Mel-Frequency Cepstral Coefficients Signal Processing," *SSRG International Journal of VLSI & Signal Processing*, vol. 3, no. 1, pp. 5-8, 2016. *Crossref*, https://doi.org/10.14445/23942584/IJVSP-V3I1P102

[19] Weili Zhou, and Zhen Zhu, “A Novel BNMF-DNN Based Speech Reconstruction Method for Speech Quality Evaluation under Complex Environments,” *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 4, pp. 959-972, 2021. *Crossref*, https://doi.org/10.1007/s13042-020-01214-3

[20] Szu-Wei Fu et al., “Quality-Net: An End-to-End Non-Intrusive Speech Quality Assessment Model Based on BLSTM,” *ArXiv preprint arXiv:1808.05344*. *Crossref*, https://doi.org/10.48550/arXiv.1808.05344

[21] Zeng Runhua, and Zhang Shuqun, "Improving Speech Emotion Recognition Method of Convolutional Neural Network," *International Journal of Recent Engineering Science*, vol. 5, no. 3, pp. 1-7, 2018. *Crossref*, https://doi.org/10.14445/23497157/IJRES-V5I3P101

[22] Jagadish S.Jakati, and Shridhar S.Kuntoji, “Novel Speech Enhancement Solution Using Hybrid Wavelet Transformation Least Means Square Method,” *International Journal of Engineering Trends and Technology*, vol. 69, no. 7, pp. 233-243, 2021. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V69I7P230

[23] Kristian Timm Andersen, and Marc Moonen, “Adaptive Time-Frequency Analysis for Noise Reduction in an Audio Filter Bank with Low Delay,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 4, pp. 784-795, 2016. *Crossref*, https://doi.org/10.1109/TASLP.2016.2526779

[24] Manjeet Singh, and Er.Naresh Kumar Garg, "Audio Noise Reduction Using Butter worth Filter," *International Journal of Computer & Organization Trends*, vol. 4, no. 2, pp. 20-23, 2014.

[25] Ying Deng, V. John Mathews, and Behrouz Farhang-Boroujeny, “Low-Delay Nonuniform Pseudo-QMF Banks with Application to Speech Enhancement,” *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 2110-2121, 2007.

[26] Karush Suri, "Sub - Band Coding and Speech Quality Testing," *SSRG International Journal of Electronics and Communication Engineering*, vol. 3, no. 1, pp. 10-13, 2016. *Crossref*, https://doi.org/10.14445/23488549/IJECE-V3I1P104

[27] Jagadish S. Jakati, and Shridhar S. Kuntoji, "A Noise Reduction Method Based on Modified LMS Algorithm of Real Time Speech Signals," *WSEAS Transactions on Environment and Development*, vol. 16, no. 13, pp. 162-170, 2021. *Crossref*, https://doi.org/10.37394/23203.2021.16.13

[28] Fairriky Rynjah, Bronson Syiem, and L. Joyprakash Singh, "Investigating Khasi Speech Recognition Systems using a Recurrent Neural Network-Based Language Model," *International Journal of Engineering Trends and Technology*, vol. 70, no. 7, pp. 269-274, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I7P227

[29] Jagadish S.Jakati, and Shridhar S.Kuntoji, "Efficient Speech De-Noising Algorithm Using Multi-Level Discrete Wavelet Transform and Thresholding," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 6, pp. 2472-2480, 2020. *Crossref*, https://doi.org/10.30534/ijeter/2020/43862020

[30] Sara Sandabad, Achraf Benba, and Hasna Nhaila, "Parkinson's Syndrome Diagnosis Applying Perceptual Linear Prediction Cepstral Analysis on Several Speech Recordings," *International Journal of Engineering Trends and Technology*, vol. 70, no. 9, pp. 214-221, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I9P222

[31] Anil Garg, and O. P. Sahu, "A Hybrid Approach for Speech Enhancement Using Bionic Wavelet Transform and Butterworth Filter," *International Journal of Computers and Applications*, vol. 42, no. 7, pp. 686-696, 2020. *Crossref*, https://doi.org/10.1080/1206212X.2019.1614293

[32] Hyeong-Seok Choi et al., "Phase-Aware Speech Enhancement with Deep Complex U-Net," *International Conference on Learning Representations*, 2019.

[33] Rafael Attili Chiea, Márcio Holsbach Costa, and Guillaume Barraultb, "New Insights on the Optimality of Parameterized Wiener Filters for Speech Enhancement Applications," *Speech Communication*, vol. 109, pp. 46-54, 2019. *Crossref*, https://doi.org/10.1016/j.specom.2019.03.005