RESEARCH ARTICLE

INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY TECHNOVATION

Asian Research Association

DOI: 10.54392/irjmt2522

# Refining Speech Clarity with Wavelet Denoising under Different Face Mask Conditions: A Subjective Analysis

**B. Marxim Rahula Bharathi [a,*], Adireddy Ramesh [b], N.S. Balaji [c], P.V. Elumalai [a] Akhilesh Kumar Singh [a], V.V.M.J. Satish Chembuly [a], Huaizhi Zhang [d]**

[a] Department of Mechanical Engineering, Aditya University, Surampalem, Andhra Pradesh, India.

[b] Department of Electrical and Electronic Engineering, Aditya University, Surampalem, Andhra Pradesh, India.

[c] Department of Mechanical Engineering, SRM Institute of Science and Technology, Tiruchirappalli Campus, Tiruchirappalli, Tamil Nadu, India.

[d] Faculty of Engineering & Technology, Shinawatra University, Bang Toei, Thailand - 12160.

* Corresponding Author Email: prof.bmrbharathi@gmail.com

DOI: https://doi.org/10.54392/irjmt2522

Received: 22-10-2024; Revised: 21-01-2025; Accepted: 11-02-2025; Published: 24-02-2025

Check for updates

**Abstract:** Amid the COVID-19 pandemic, people have adopted various face masks and face shields as protective measures against infection. While these measures have been instrumental in saving countless lives, they pose significant challenges to interpersonal communication, especially in scenarios requiring clear verbal interaction. This study utilizes a microphone to capture speech signals in different scenarios involving face masks, with and without face shields. Participants read vowels and the Grandfather Passage across ten experimental conditions, including surgical masks, cloth masks, double masks (surgical and cloth combination), and N95 masks, both with and without face shields. The obtained speech signals, often distorted by noise and reverberation, undergo enhancement through the wavelet denoising approach using discrete wavelet transform with soft thresholding. The quality of the enhanced signals was compared to the original acquired signals using a subjective comparison test involving 30 listeners who rated the signals based on comparison mean opinion scores (CMOS). Multiple research findings indicate that the signal improvement achieved through wavelet denoising consistently exceeds the quality of the initial signal, even under challenging conditions such as double masks with face shields. This study highlights the practical efficacy of wavelet denoising in addressing speech clarity challenges caused by protective face coverings, offering a valuable solution for improved communication in masked environments.

**Keywords:** Face masks, Speech Enhancement, Wavelet Transform, Subjective comparison test, Wavelet Denoising, Process Innovation

## 1. Introduction

Face masks and shields play an important role in preventing the spread of the Coronavirus outbreak [1, 2]. Millions of individuals have been protected as a result. During the pandemic, it is unavoidable for people to wear face masks in everyday communication situations. Consequently, these face masks and shields have hindered people's conversations [3-6]. Some individuals experience difficulties understanding others' words when wearing masks and face shields. Speech enhancement processing can assist in resolving this issue. One of the classic signal processing problems is speech enhancement. Palmiero et al. [3] and Radonovich et al. [7] investigated the impact of speech intelligibility while using respiratory masks.

Corey, et al. [5] discovered from various experiments that face masks attenuate high-frequency. In their study, Balamurali et al. [8] examined the impact of design and material selection on the frequency response of different face masks. They also explored how these masks could affect the production and transmission of speech sounds at the speaker's lips. Their findings suggested that material selection has a greater influence on transmission than mask geometry and design would imply.

The process of speech signal enhancement aims to enrich the clarity, intelligibility, and understanding of acquired speech signals. There are various methods for enhancing speech signals, which can be categorized as single and multi-channel [9, 10]. Single-channel systems are well-suited for real-time applications such as individual hearing aids and smart classroom audio setups. Conversely, multi-channel audio enhancement is employed in systems that handle multiple speech signal inputs. The spectral subtraction method involves using the Fourier converts to convert the time domain speech signal into the frequency domain speech signal. This method was proposed by Nelson to eliminate background noise [11].

Different techniques, including removing recurring noise, and interfering acoustic speech techniques, can be applied to remove noise from speech signals [12]. F. Qu et al. improved speech signals using an adaptive smoothing factor-based spectral subtraction method [13], while Paliwal K et al. proposed a method based on short-time modulation to minimize musical noise [14]. Hu et al. significantly improved speech signal quality using a special subtraction method with spectral smoothing and comb filtering [15]. Additionally, the adaptive spectral subtraction method for noise cancellation, which requires more than one microphone, was also proposed [16].

Wavelet analysis and wavelet denoising have become significant research aspects in various fields, including object localization and tracking [17-19], noise reduction in hearing aids [20], bio-acoustics [21, 22], rotating machinery fault diagnosis [23], and image processing [24] including image splicing forgery detection [25]. Over the last two decades, wavelet denoising-based speech enhancement has become a significant process in improving speech signal quality. Bahoura and Rouat proposed wavelet transform as an alternative to Short Time Fourier Transform (STFT) for representing localized events [26]. The scale and translation function similarly to STFT. Bhowmick and Chandra established a speech enhancement system using wavelet decomposition based on voiced speech probability (VSP) [27]. Ghanbari et al. proposed an adaptive threshold speech enhancement method for wavelet packets [28]. The technique was tested on various speakers and noise situations, showing significant enhancement in both signal-to-noise ratio (SNR) and automatic speech recognition (ASR) results. Marxim et al. aimed to enhance speech signal quality degraded by face masks using the Empirical Mode Decomposition (EMD) denoising method. This approach decomposes speech signals into intrinsic mode functions, isolates noise, and enhances the acoustic signal. Their results demonstrated that the EMD method outperformed the traditional spectral subtraction technique in improving speech clarity under mask-induced conditions, highlighting its effectiveness in challenging acoustic environments [29, 30].

Senthamizh Selvi R et al proposed a novel approach to speech enhancement merges the Discrete Wavelet Transform (DWT) and Long Short-Term Memory (LSTM) algorithms. DWT decomposes noisy speech signals into frequency components, while LSTM networks enhance the denoised signals, thereby elevating speech quality. This methodology employs DWT to decompose noisy speech signals into frequency components, followed by LSTM networks for denoising. The resulting speech quality and intelligibility are notably enhanced, attaining an output Signal-to-Noise Ratio (SNR) of 30.6dB at an input noise level of 15dB [31]. ZT Wu et al. proposed a speech enhancement approach using FullSubNet+ and adaptive-FSN, neural network models that integrate full-band and sub-band fusion. By replacing the traditional complex spectrogram with discrete wavelet transform (DWT) features, they improved phase information representation. This enhancement led to higher speech quality and intelligibility compared to the original adaptive-FSN, showcasing the effectiveness of combining wavelet transforms with neural networks for speech enhancement [32].

The highlights of the research work are listed below:

- Application of Wavelet Denoising Speech Enhancement Technique: The research effectively applies the Wavelet Denoising Speech Enhancement to improve single-channel speech input signal across various scenarios, including different face masks and face shield conditions.
- Verification Using Comparison Mean Opinion Scores (CMOS): The efficiency of Wavelet-Denoising speech signal enhancement undergoes thorough validation. A comprehensive subjective assessment was carried out utilizing CMOS 30 participants. These assessments provide valuable insights into the enhanced speech quality achieved through the proposed technique.

The wavelet denoising method is employed in this study to enhance the speech signal in various face masks and face shield circumstances. Section 2 describes the experimental setup and execution, while Section 3 explains the wavelet denoising speech enhancement. Section 4 discusses the experiments, presents the findings, and includes the assessment of speech quality. Section 5 summarizes the findings.

## 2. Methods

Figure 1 shows the schematic representation of the research methods used in this research work. Speech signals were obtained from four participants, all of whom, despite having English as their medium of instruction, are non-native English speakers. The investigation incorporates various face mask conditions, including those with and without a face shield. The conditions include no mask, surgical mask, cloth mask, double masks (a combination of surgical and cloth masks), and N95 mask, as detailed in Table 1. The experimental setup is depicted in Figure 2, illustrating the different face mask and face shield conditions during the experiments.

As outlined in Table 1, all participants were instructed to read vowels and the grandfather passage (GFP) [33] under varied mask conditions, both with and without face protection. The methodology involved repeating the process five times for vowels and three times for GFP. Speech signals were acquired by a microphone at a sampling rate of 44 kHz and a resolution of 16-bit. The recorded data was stored in WAV format on a personal computer.

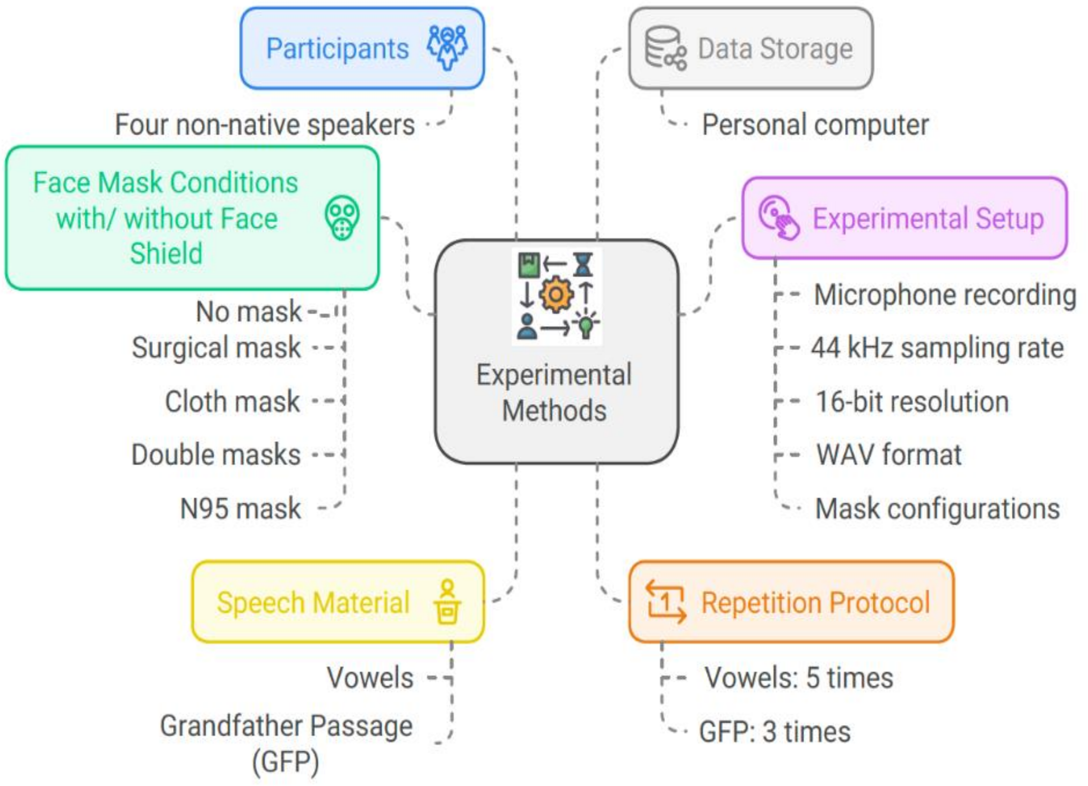
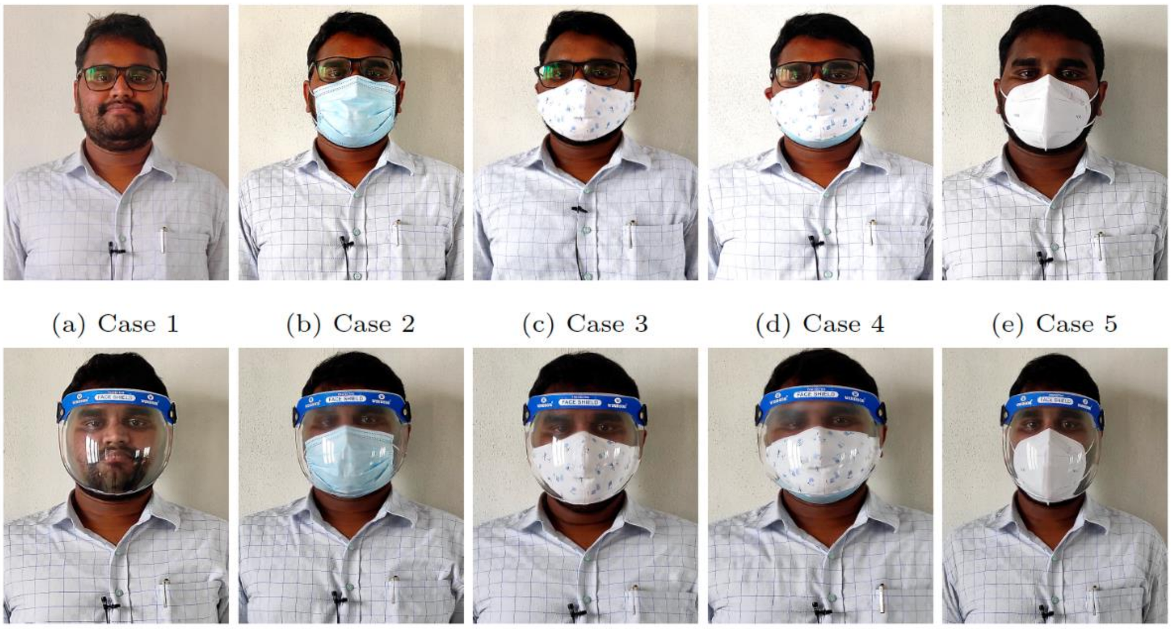



**Figure 1.** Schematic Diagram of Experimental Methods



(a) Case 1 (b) Case 2 (c) Case 3 (d) Case 4 (e) Case 5
(f) Case 6 (g) Case 7 (h) Case 8 (i) Case 9 (j) Case 10

**Figure 2.** Face mask and face shield conditions

**Table 1.** Face Mask Conditions for Experiments

| CASE NUMBER | CONDITION |
|---|---|
| Case - 1 | Without face shield and without mask |
| Case - 2 | Without face shield and Surgical mask |
| Case - 3 | Without face shield and Cloth mask |
| Case - 4 | Without face shield and Double mask |
| Case - 5 | Without face shield and N95 mask |
| Case - 6 | Face shield and without mask |
| Case - 7 | Face shield and Surgical mask |
| Case - 8 | Face shield and Cloth mask |
| Case - 9 | Face shield and Double mask |
| Case - 10 | Face shield and N95 mask |

## 3. Wavelet Denoising - Speech Signal Enhancement

The wavelet transform (WT) is one of the key methods used in digital signal processing for time-frequency analysis. Its effective time-frequency analysis features make wavelets and wavelet denoising valuable tools in diverse research areas. Wavelets work by breaking down a signal into a linear combination of wavelet coefficient products and the mother wavelet [18]. The generalised formula for discrete wavelet transform (DWT) is given as [18],

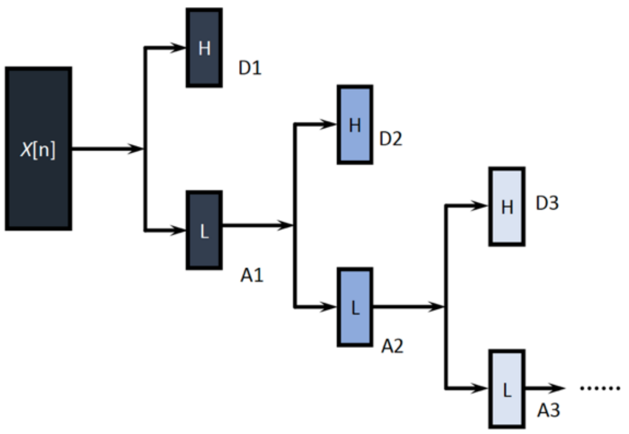



**Figure 3.** DWT decomposition

$$X[a,b] \;=\; \sum_{n=-\infty}^{\infty} x[n]\,\emptyset_{a,b}[n] \qquad (1)$$

Here, the input signal $(x[n])$ is to be transformed into the wavelet domain. The function ϕ[n] represents a finite-length window, with (a) and (b) serving as the parameters for scaling (either dilation or contraction) and shifting (translation), respectively. The function $(\phi_{a,b}[n])$ is defined as:

$$\phi_{a,b}[n] \;=\; \frac{1}{\sqrt{a}}\emptyset\left[\frac{n-b}{a}\right] \qquad (2)$$

In this expression:

- $(\phi_{a,b}[n])$ is the scaled and translated version of the window function $(\phi[n])$.
- $(a)$ is the dilation or contraction parameter, affecting the width of the window.
- $(b)$ is the translation parameter, determining the position of the window
- The sum is taken over all values of $(n)$ in the range $(-1)$ to $(+\infty)$. This mathematical formulation is a common representation of the continuous wavelet transform (CWT), where the $x[n]$ is analyzed at different scales and positions using the window function $(\phi_{a,b}[n])$.

Any function can be broken down into different scales and resolutions using the wavelet transform (WT), as shown in Figure 3. The low pass filter (L) is utilized to obtain the scaling function, while the high pass filter (H) is used to derive the wavelet function. Initially, the input signal is decomposed into D1 and A1, where D1 represents the detail component and A1 represents the approximation component at level one. Once D1 and A1 undergo processing through highpass and lowpass filters, they generate second-level components D2 and A2. This iterative process continues until the desired level of decomposition is achieved.

## 4. Results and Discussions

The microphone captures the speech signal under various testing conditions as specified in Section 2. In certain experimental scenarios involving the use of a face shield, the spoken signal may include echoes. Consequently, these temporal domain signals comprise both noise and reverberation. The reverberation time was measured across a frequency range of 250 Hz to 8000 Hz, revealing a trend of decreasing reverberation time with increasing frequency.

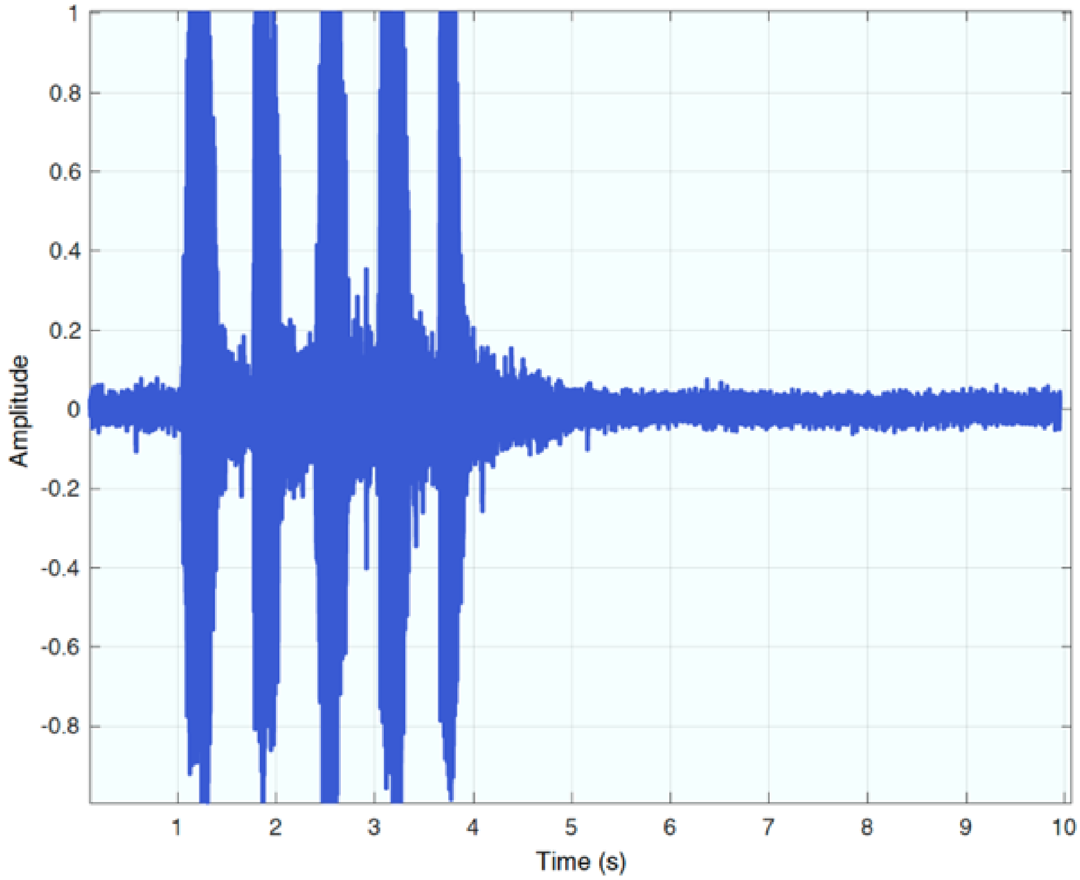



**Figure 4.** Case 9: Acquired Time Domain Signal.

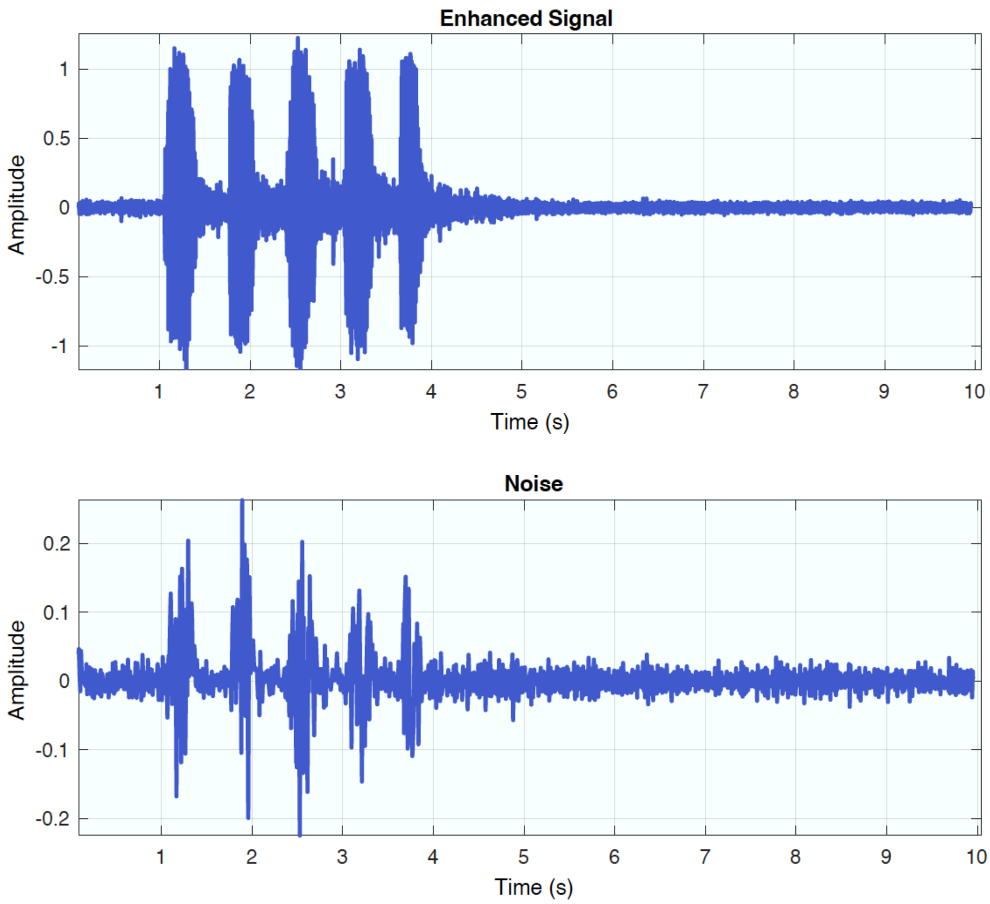



**Figure 5.** Case 9: Wavelet Denoising Speech Enhancement for vowels (a) Enhanced Signal (b) Noise

Specifically, the measured values were 0.952 seconds at 250 Hz, 0.878 seconds at 500 Hz, 0.821 seconds at 1000 Hz, 0.776 seconds at 2000 Hz, and 0.642 seconds at 4000 Hz [29]. The analysis of a speech signal becomes challenging when it incorporates such unwanted elements. Participants were instructed to read both vowels and the grandfather passage (GFP) in all ten experimental conditions, as outlined in Section 2. The microphone picks up these acoustic waves, and the signals are subsequently recorded in the computer.

Figure 2 shows the time-domain representation of the speech signal recorded during vowel articulation under case 9 conditions, where two face masks and a face shield were used.The double mask, comprising a surgical mask and a cloth mask, leads to notable signal attenuation. The acquired signal undergoes decomposition into various levels using discrete wavelet transform. For both male and female speakers, the soft threshold and mother wavelet Db10 are selected [28]. In Figure 3 (a), the wavelet denoising-enhanced speech signal for case 9 during vowel reading is presented, while Figure 3 (b) shows the corresponding noise signal. Similarly, Figure 4 (a) displays the wavelet denoising-enhanced speech signal for case 9 during reading GFP, and Figure 4 (b) presents the noise signal for the same case. Quantifying the speech signal enhancement in the time domain based on Figure 3 (a) and Figure 4 (a) is challenging. Therefore, a subjective listening test is employed to evaluate wavelet denoising-based speech enhancement. Thirty listeners assessed the enhancement of speech quality using CMOS [34]. Out of the thirty listeners, 77% are male and 23% are female. In terms of age, 37% of the listeners are between 21 and 30, 23% are over 50, 20% are under 20, 13% are between 31 and 40, and 7% are between 41 and 50.

The acquired raw signal and the wavelet-enhanced vowel signal are played for the listeners. They are instructed to evaluate the second signal by comparing it to the first signal and assigning a rating on a scale from 3 to -3. Table 2 outlines seven phases of the comparative category rating used in this evaluation. The identical procedure is followed for the GFP recorded speech signal. The acquired speech signal that was obtained and the speech signal that has been improved through wavelet denoising are referred to as signal one and signal two, respectively. The subjective comparison test results for the original raw signal and the wavelet-denoised speech-enhanced signal are shown in Figure 5, focusing on vowels. According to Figure 5, in case one (without face mask and face shield), 16 listeners rate signal two as better than signal one, and six listeners rate signal two as much better. Additionally, eight listeners perceive signal two as similar to signal one. Overall, all listeners concur that the enhanced signal is an improvement over the acquired signal. In cases two to five, the majority of listeners expect signal two to be superior to signal one.

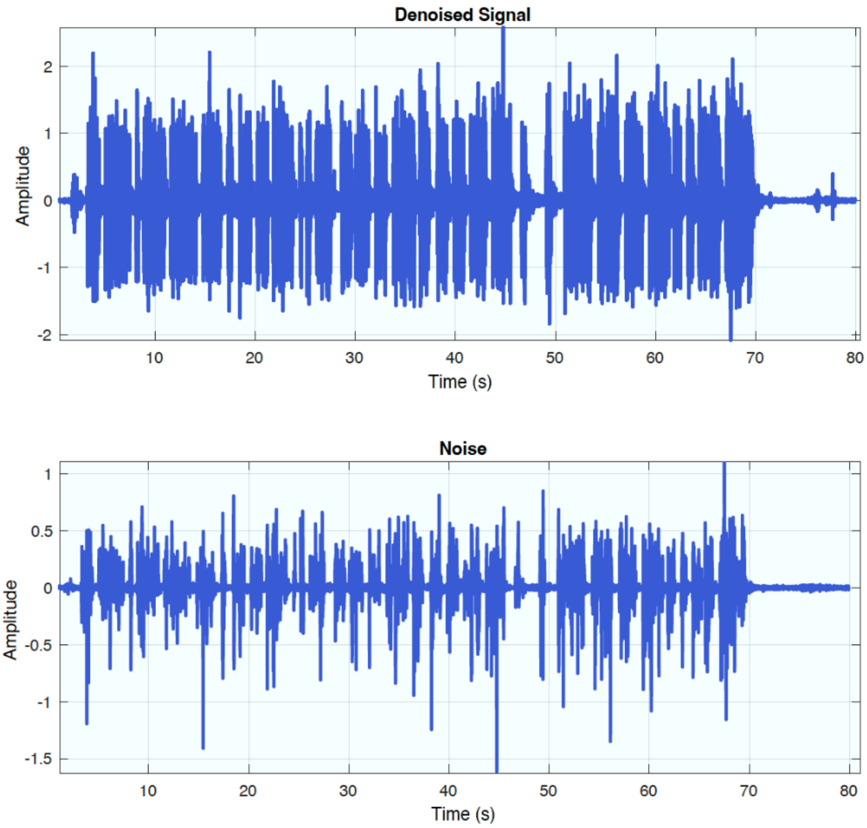



**Figure 6.** Case 9: Wavelet Denoising Speech Enhancement for GFP (a) Enhanced Signal (b) Noise

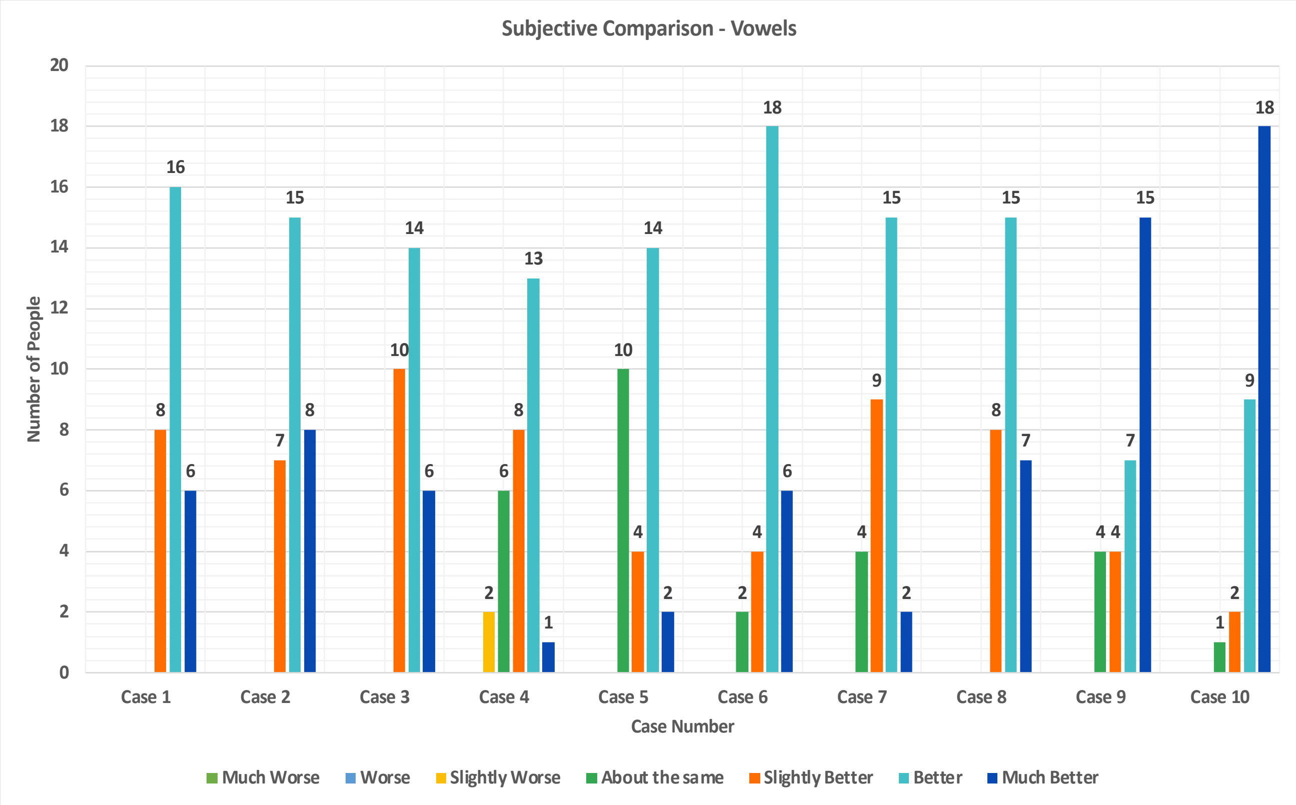



**Figure 7.** Results of Subjective Comparison – Vowels

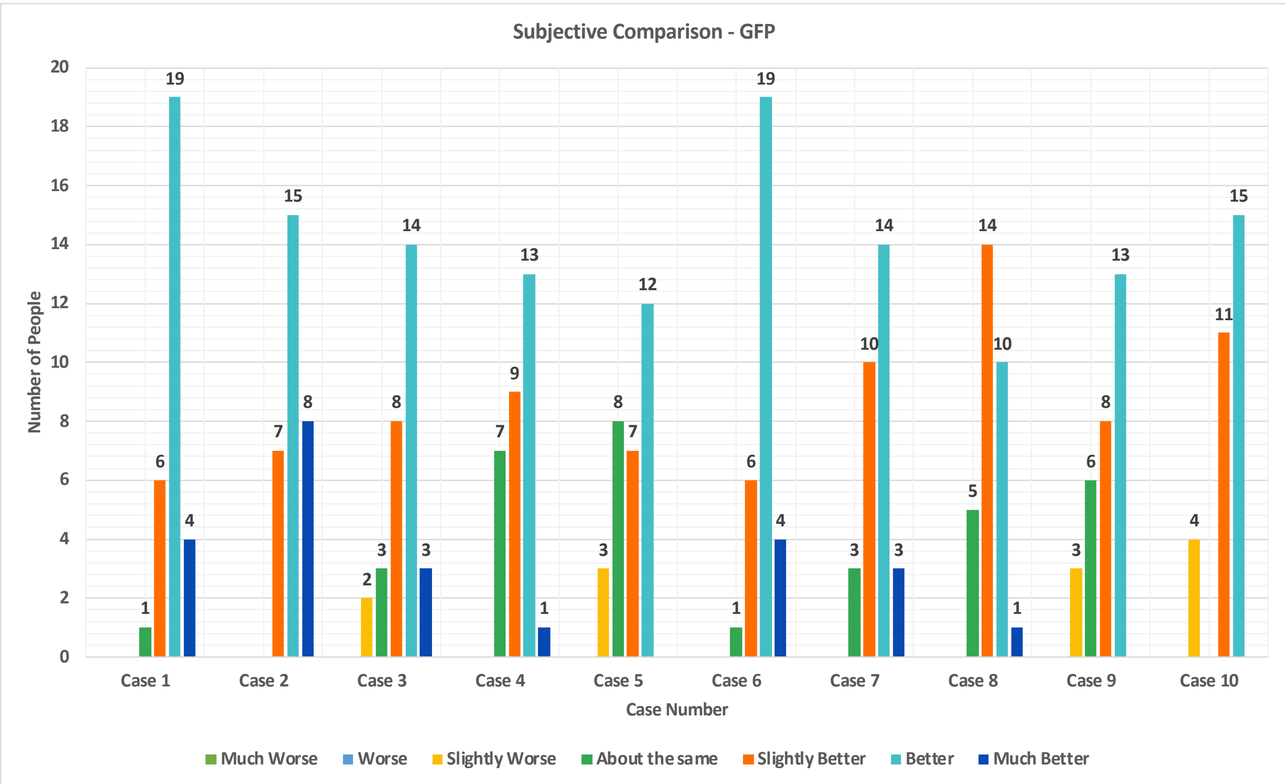



**Figure 8.** Results of Subjective Comparison - GFP

Similarly, according to Figure 5, in case six, 18 listeners rate signal two as better than signal one, and six listeners rate signal two as much better. Furthermore, four listeners rate signal two as slightly better, and two listeners perceive signal two as similar to signal one. Overall, all listeners agree that the enhanced signal is an improvement over the acquired signal. Except for case eight, where the majority of listeners anticipate signal two to be better than signal one in cases seven to ten. In case eight (double face mask and face shield condition), the acquired speech signal exhibits high noise and reverberation compared to other conditions. Despite this, more than 73 percent of listeners affirm that the wavelet denoising speech enhancement process is effective.

The subjective comparison test results for the acquired raw signal and the wavelet denoised speech-enhanced signal are depicted in Figure 6 specifically for GFP. According to Figure 6, in case one (without face mask and face shield), 19 listeners rate signal two as better than signal one, and four listeners rate signal two as much better. Additionally, six listeners perceive signal two as slightly better, and one listener agrees that signal two is similar to signal one. Overall, all listeners assert that the enhanced signal is an improvement over the acquired signal. In cases two to five, the majority of listeners expect signal two to be superior to signal one. Similarly, according to Figure 6, in case six, 19 listeners rate signal two as better than signal one, and four listeners rate signal two as much better. Furthermore, six listeners rate signal two as slightly better, and one listener agrees that signal two is similar to signal one.

The Segmental Signal-to-Noise Ratio, commonly referred to as Segmental SNR (SSNR), is a method used to evaluate speech quality by calculating the average SNR over brief, fixed-length segments of the signal, typically spanning 15–20 milliseconds. Unlike traditional SNR calculations that assess the entire signal, this localized approach enables a more detailed analysis of variations in signal clarity across different temporal frames. The Segmental SNR is derived using the equation [35, 36].

$$[\mathrm{SNR}_{\mathrm{seg}}(\mathrm{dB}) = \frac{1}{M}\sum_{m=1}^{M} 10\log_{10}\left(\frac{\sum_{n=1}^{L} x^2(n)}{\sum_{n=1}^{L}[x(n)-x_s(n)]^2}\right)] \quad (3)$$

where $x(n)$ represents the clean speech signal, $x_s(n)$ is the enhanced or processed signal, $M$ denotes the total number of frames in the signal, and $L$ is the frame length. This metric captures the energy of the clean speech signal relative to the error introduced by the enhancement process, averaged across all frames. By focusing on individual segments rather than the entire signal, Segmental SNR provides a rigorous and precise evaluation of the performance of speech enhancement techniques, highlighting their impact on noise suppression and speech intelligibility.

The SSNR method was employed as an objective evaluation framework to quantify speech quality across various face mask and face shield conditions, as shown in Figure 9. The analysis revealed that cases 10, 9, and 6 performed optimally for vowels, indicating substantial suppression of noise and enhancement of speech clarity in these scenarios. Similarly, for GFP, cases 1, 6, and 10 demonstrated the highest Segmental SNR values, signifying improved intelligibility under these conditions. Comparative analysis with the subjective evaluation based on CMOS revealed a notable alignment in key cases, particularly for case 6 and case 10, both of which consistently performed well in both frameworks.

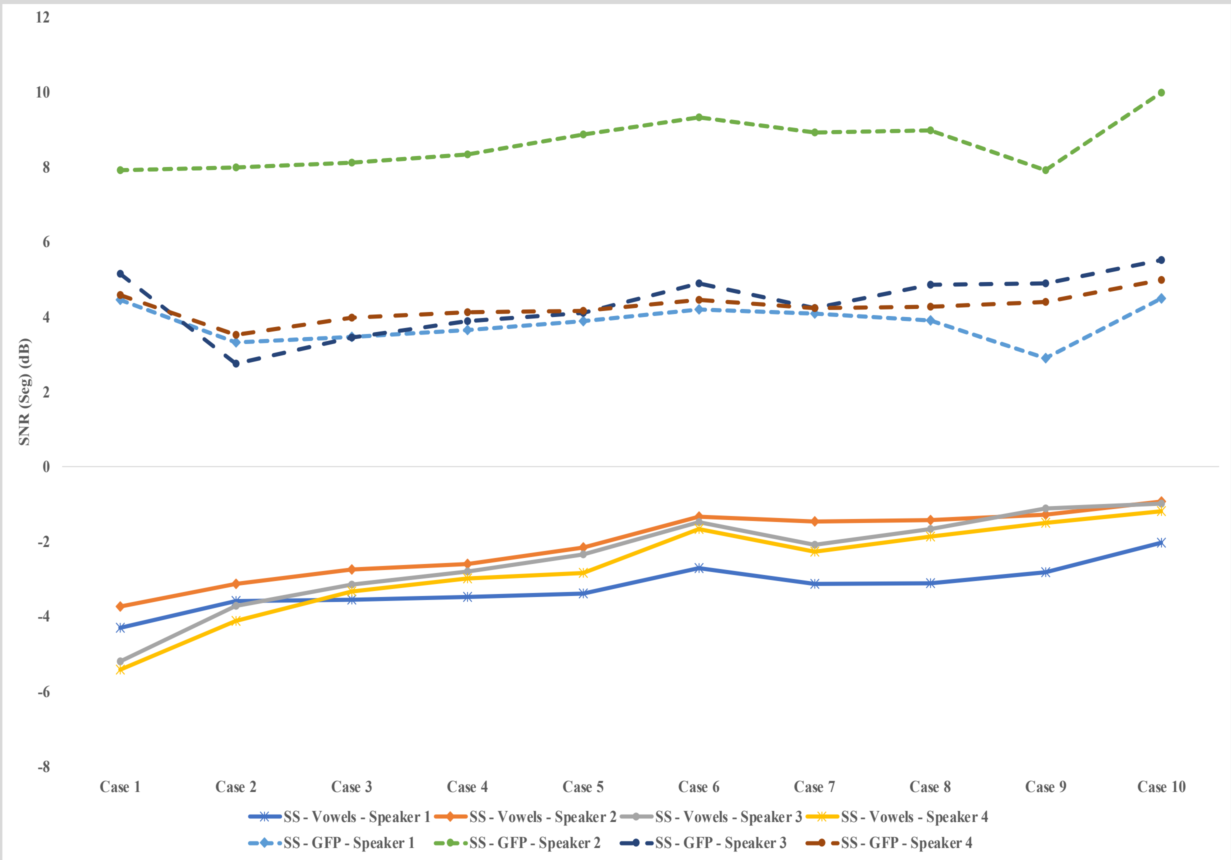



**Figure 9.** Results of Segmental SNR for Vowels and GFP

This concurrence underscores the reliability of the enhancement technique in mitigating speech degradation caused by noise and reverberation. However, discrepancies were observed in specific scenarios, such as GFP case 1, where the objective SSNR indicated strong performance, but subjective CMOS ratings reflected more variability. Such differences may stem from the perceptual nuances of human evaluators, emphasizing the necessity of integrating both objective and subjective analyses to achieve a holistic assessment of speech quality. This comparative study validates the robustness of wavelet denoising in diverse acoustic environments while highlighting areas for further investigation.

Overall, all listeners concur that the enhanced signal is an improvement over the acquired signal. Except for case eight, where the majority of listeners anticipate signal two to be better than signal one in cases seven to ten. In case eight (double face mask and face shield condition), the acquired speech signal exhibits high noise and reverberation compared to other conditions. Despite this, more than 83 percent of listeners agree that the wavelet denoising speech enhancement process is effective. These results demonstrate that wavelet denoising-based speech enhancement is beneficial for improving speech signals in various face masks and face shield conditions.

## 5. Conclusions

The vowels and the Grandfather Passage (GFP) were read by participants while wearing various face masks and face shields, and the corresponding speech signals were recorded using a microphone. These signals, often distorted by noise and reverberation, were decomposed at different resolutions using the wavelet transform, followed by wavelet denoising to enhance the speech quality. A comparison of the raw signal and the wavelet-denoised signal was conducted with the participation of thirty individuals. The results demonstrated that the wavelet denoising method consistently improved the clarity and quality of speech across a wide range of face mask conditions, including those with and without face shields. More than 73% of listeners rated the enhanced signal as better than the raw signal in challenging scenarios, such as double masks with face shields. Future research could incorporate wavelet denoising into real-time communication systems, which could offer significant advantages in environments such as healthcare, education, telecommunication, and customer service, where clear communication is critical.

## References


[1] J. Brainard, N. Jones, I. Lake, L. Hooper, P.R. Hunter, Facemasks and similar barriers to prevent respiratory illness such as covid-19: A rapid systematic review. MedRxiv, (2020) 2020 - 04. https://doi.org/10.1101/2020.04.01.20049528

[2] C. J. Worby, H. H. Chang, Face mask use in the general population and optimal resource allocation during the covid-19 pandemic. Nature communications, 11(1), (2020), 1-9. https://doi.org/10.1038/s41467-020-17922-x

[3] A.J. Palmiero, D. Symons, J.W. Morgan III, R.E. Shaffer, Speech intelligibility assessment of protective facemasks and air-purifying respirators. Journal of occupational and environmental hygiene, 13(12), (2016), 960 – 968. https://doi.org/10.1080/15459624.2016.1200723

[4] S.R. Atcherson, L.L. Mendel, W.J. Baltimore, C. Patro, S. Lee, M. Pousson, M.J. Spann, The effect of conventional and transparent surgical masks on speech understanding in individuals with and without hearing loss. Journal of the American Academy of Audiology, 28(01), (2017) 058 - 067. https://doi.org/10.3766/jaaa.15151

[5] R.M. Corey, U. Jones, A.C. Singer, Acoustic effects of medical, cloth, and transparent face masks on speech signals. The Journal of the Acoustical Society of America, 148(4), (2020) 2371 - 2375. https://doi.org/10.1121/10.0002279

[6] B. Gursharan, T. Knowles, Acoustic and perceptual impact of face masks on speech: A scoping review. Plos one, 18(8), (2023) e0285009. https://doi.org/10.1371/journal.pone.0285009

[7] L.J. Radonovich R. Yanke, J. Cheng, B. Bender, Diminished speech intelligibility associated with certain types of respirators worn by healthcare workers. Journal of occupational and environmental hygiene, 7(1), (2009) 63 - 70. https://doi.org/10.1080/15459620903404803

[8] B. Balamurali, T. Enyi, C.J. Clarke, S.Y. Harn, J.M. Chen, Acoustic effect of face mask design and material choice. Acoustics Australia, 49(3), (2021) 505 – 512. https://doi.org/10.1007/s40857-021-00245-2

[9] D.S. Kulkarni, R.R. Deshmukh, P.P. Shrishrimal, A review of speech signal enhancement techniques. International Journal of Computer Applications, 139(14), (2016) 23-26. https://doi.org/10.5120/ijca2016909507

[10] N.V. Lalitha, G. Suresh, M.K. Singh, V. M. Kumar, Speech Signal Splicing Detection System Based on MFCC and DTW. International Research Journal of Multidisciplinary Technovation 6 (6), (2024)186-97. https://doi.org/10.54392/irjmt24613

[11] D.J. Nelson, (2005) Signal reconstruction from concentrated STFT peaks [short time Fourier transform peaks]. Proceedings. (ICASSP'05). IEEE International Conference on Acoustics,

Speech, and Signal Processing, IEEE, USA. https://doi.org/10.1109/ICASSP.2005.1415962

[12] A. Chaudhari, S.B. Dhonde, (2015) A review on speech enhancement techniques. International Conference on Pervasive Computing (ICPC), IEEE, India. https://doi.org/10.1109/PERVASIVE.2015.7087096

[13] F. Qu, S. Lei, Z. Zhao, J. Zhang, Z. Nie, (2021) A modified a priori SNR estimation for spectral subtraction speech enhancement. In 2021 IEEE 4th International Conference on Electronics Technology (ICET), IEEE, China. https://doi.org/10.1109/ICET51757.2021.9451018

[14] K. Paliwal, K. Wójcicki, B. Schwerin, Single-channel speech enhancement using spectral subtraction in the short-time modulation domain. Speech communication, 52(5), (2010) 450-475. https://doi.org/10.1016/j.specom.2010.02.004

[15] H.T. Hu, F.J. Kuo, H.J. Wang, Supplementary schemes to spectral subtraction for speech enhancement. Speech Communication, 36(3-4), (2002) 205-218. https://doi.org/10.1016/S0167-6393(00)00086-8

[16] M.F. Kasim, T. Adiono, M. Fahreza, M.F. Zakiy, Real-time architecture and FPGA implementation of adaptive general spectral substraction method. Procedia Technology, 11, (2013) 191-198. https://doi.org/10.1016/j.protcy.2013.12.180

[17] O. Postolache, P. Girao, M. Pereira. Underwater acoustic source localization based on passive sonar and intelligent processing. IEEE Instrumentation & Measurement Technology Conference IMTC, IEEE, Poland. https://doi.org/10.1109/IMTC.2007.379152

[18] M.R.B. Boopathi Rajan, A.R. Mohanty, Time delay estimation using wavelet denoising maximum likelihood method for underwater reverberant environment. IET Radar, Sonar & Navigation, 14(8), (2020) 1183-1191. https://doi.org/10.1049/iet-rsn.2020.0079

[19] S. Zhang, S. Wan, Y. Wang, B. Zhang, Z. Zhang, H. Zhong, J. Shi, J. Sun, X. He, Q. Wu, 2D sound source localization technology based on diaphragm EFPI fiber microphone array. Optics Communications, 519, (2022) 128435. https://doi.org/10.1016/j.optcom.2022.128435

[20] Whitmal, Nathaniel A., and Janet C. Rutledge. Noise reduction in hearing aids: a case for wavelet-based methods. Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No. 98CH36286). Vol. 3. IEEE, 1998. https://doi.org/10.1109/IEMBS.1998.747070

[21] M.B. Gur, C. Niezrecki, A wavelet packet adaptive filtering algorithm for enhancing manatee vocalizations. The Journal of the Acoustical Society of America, 129(4), (2011) 2059-2067. https://doi.org/10.1049/iet-rsn.2020.0079

[22] K. Leftwich, J.W. Ioup, Denoising and deconvolving sperm whale data in the northern Gulf of Mexico. The Journal of the Acoustical Society of America, 151(4_Supplement), (2022) A136-A136. https://doi.org/10.1121/10.0010899

[23] C. Beale, C. Niezrecki, M. Inalpolat, An adaptive wavelet packet denoising algorithm for enhanced active acoustic damage detection from wind turbine blades. Mechanical Systems and Signal Processing, 142, (2020) 106754. https://doi.org/10.1016/j.ymssp.2020.106754

[24] S.P. Vaidya, P.V.S.S. Mouli, Robust digital color image watermarking based on compressive sensing and DWT. Multimedia Tools and Applications, 83, (2024) 3357–3371. https://doi.org/10.1007/s11042-023-15349-2

[25] M.K. Singh, DWT and LBP hybrid feature based deep learning technique for image splicing forgery detection. Soft Computing, 28, (2024) 12207–12215. https://doi.org/10.1007/s00500-024-09919-1

[26] M. Bahoura, J. Rouat, Wavelet speech enhancement based on time–scale adaptation. Speech Communication, 48(12), (2006) 1620-1637. https://doi.org/10.1016/j.specom.2006.06.004

[27] A. Bhowmick, M. Chandra, Speech enhancement using voiced speech probability based wavelet decomposition. Computers & Electrical Engineering, 62, (2017) 706-718. https://doi.org/10.1016/j.compeleceng.2017.01.013

[28] Y. Ghanbari, M.R. Karami-Mollaei, A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets. Speech communication, 48(8), (2006) 927-940. https://doi.org/10.1016/j.specom.2005.12.002

[29] M.R. Bharathi, N.S. Balaji, A.K. Singh, R. Sundaramurthi, Improving speech communication in the age of face masks: A study on EMD denoising method by subjective speech comparison. e-Prime-Advances in Electrical Engineering, Electronics and Energy, 5, (2023) 100267. https://doi.org/10.1016/j.prime.2023.100267

[30] S.R. Senthamizh, G.G. Naidu, G. Tejaswini, (2024) Speech Enhancement using Discrete Wavelet Transform with Long Short-Term Memory Algorithm. Nanotechnology Perceptions, Brookfield Academic Limited, United Kingdom.

[31] Z.T. Wu, P.F. Li, P.C. Wu, E.S. Li, J.W. Hung, (2023). Exploiting Discrete Wavelet Transform Features in Speech Enhancement Technique

Adaptive FullSubNet+. In 2023 International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan), IEEE, Taiwan. https://doi.org/10.1109/ICCE-Taiwan58799.2023.10226809

[32] J. Reilly, J.L. Fisher, Sherlock Holmes and the strange case of the missing attribution: A historical note on "The Grandfather Passage. Journal of Speech, Language, and Hearing Research, 55(1), (2012) 84-88. https://doi.org/10.1044/1092-4388(2011/11-0158)

[33] X. Tan, J. Chen, H. Liu, J. Cong, C. Zhang, Y. Liu, X. Wang, Y. Leng, L. He, S. Zhao, T. Qin, F. Soong, T.Y. Liu, Naturalspeech: End-to-end text-to-speech synthesis with human-level quality. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(6), (2024) 4234 - 4245.

[34] R.K. Kandagatla, V.J. Naidu, P.S. Reddy, M. Gayathri, A. Jahnavi, K. Rajeswari, Analysis of statistical estimators and neural network approaches for speech enhancement. Science and Engineering Journal,17(Supplement), (2024) 17-27. https://doi.org/10.54645/202417SupXBB-31

[35] K. Taira, K. Kondo, (2015) Estimation of binaural intelligibility using the frequency-weighted segmental SNR of stereo channel signals, Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Hong Kong, China. https://doi.org/10.1109/APSIPA.2015.7415459

[36] P. Selvaraj, S.S. Maidin, & Q. Yang, (2025). Speech Enhancement using Sliding Window Empirical Mode Decomposition with Median Filtering Technique. Journal of Applied Data Sciences, 6(1), 143-154. https://doi.org/10.47738/jads.v6i1.470

## Acknowledgement

I would like to extend my heartfelt gratitude to Mr. Tarun, Mr. George, Ms. Sridevi, and Ms Meena R for their invaluable assistance during the data collection and evaluation process of this research work.

## Authors Contribution Statement

Marxim Rahula Bharathi B – Conceptualization, data collection, data Analysis and Guidance, P.V Elumalai - Writing – review & editing. N.S Balaji -Writing – review & editing, Akhilesh Kumar Singh - Data collection. Adireddy Ramesh - Data collection and analysis, VVMJ Satish Chembuly - Data collection and analysis, Huaizhi Zhang - Data analysis, Writing – review & editing. All the author's read and approved the final version of the manuscript.

## Has this article screened for similarity?

Yes

## Funding

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

## Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

## Data Availability

The data supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

## About the License