iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VII **Month of publication:** July 2026

DOI: https://doi.org/10.22214/ijraset.2026.84412

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 14 Issue VII July 2026- Available at www.ijraset.com

Acoustic Echo Cancellation Using NLMS Adaptive Filtering and Modulation-Domain Suppression

Lakumalla Pallavi[1], Dr. M. Madhavi Latha[2]
M.Tech in Systems and Signal Processing, Department of Electronics and Communication Engineering, JNTUH University College of Engineering, Science & Technology Hyderabad, Telangana, India

Abstract: ***In hands-free communication systems, acoustic echo considerably reduces speech quality and intelligibility, especially in reverberant environments. A hybrid acoustic echo cancellation system that combines modulation-domain speech augmentation, Normalized Least Mean Square (NLMS) adaptive filtering, and Room Impulse Response (RIR) simulation is presented in this research. The Image Source Method is used to create realistic acoustic environments and produce RIRs for three different reverberation situations. While STFT-based Wiener filtering minimizes residual echo and background noise while maintaining the intended speech signal, the NLMS adaptive filter suppresses the dominating acoustic echo. Using ERLE, SNRout, SI-SDR, and STOI, the suggested framework is assessed under six typical background noise circumstances with input SNR levels ranging from −10 dB to 20 dB. Echo suppression, voice quality, and intelligibility all consistently increase in various acoustic situations, according to experimental data.***

Keywords: Modulation-Domain Speech Enhancement, Room Impulse Response, Acoustic Echo Cancellation, NLMS Adaptive Filter, and Speech Enhancement.

I. INTRODUCTION

Voice assistants, smart speakers, and video conferencing platforms are examples of hands-free communication tools that have become essential to contemporary communication. These systems produce acoustic echo, which deteriorates speech quality and intelligibility, especially in reverberant environments, as the loudspeaker output is reflected by the surrounding environment and picked up by the microphone [5].

Because of its low computational complexity, consistent convergence, and suitability for real-time implementation, the Normalized Least Mean Square (NLMS) algorithm is frequently used for adaptive filtering in Acoustic Echo Cancellation (AEC) [3]– [5]. However, following adaptive filtering, background noise and residual echo frequently persist, reducing the overall effectiveness of speech improvement.The suggested system combines modulation-domain speech enhancement, NLMS adaptive filtering, and Room Impulse Response (RIR) simulation to get over this restriction. After NLMS-based echo cancellation, residual echo and background noise are further suppressed by modulation-domain processing [1], [2], [6]– [10]. The Image Source Method simulates realistic acoustic environments. Six realistic background noise situations, three simulated room environments (RT60 = 0.3 s, 0.6 s, and 0.9 s), and input SNR levels ranging from −10 dB to 20 dB are used to evaluate the framework. ERLE, Output SNR (SNRout), SI-SDR, and STOI are used to evaluate performance, showing steady gains in intelligibility, speech quality, and acoustic echo suppression.

II. LITERATURE REVIEW

In order to enhance speech quality and comprehensibility in hands-free communication devices, acoustic echo cancellation has been extensively studied. For acoustic echo estimation, adaptive filtering methods like LMS, NLMS, and RLS are frequently employed. NLMS is particularly popular due to its low computational complexity, consistent convergence, and applicability for real-time applications [3]– [5]. Furthermore, in order to evaluate acoustic echo cancellation algorithms, realistic reverberant settings are simulated using Room Impulse Response (RIR) modeling based on the Image Source Method [2].

Under reverberant situations, residual echo and background noise frequently persist even while adaptive filtering successfully reduces the primary acoustic echo. Therefore, to further reduce residual echo while maintaining voice quality, modulation-domain speech augmentation based on STFT and Wiener filtering has been used [1], [6]– [10].

Despite these advancements, the majority of current research assesses modulation-domain enhancement and adaptive filtering independently or in restricted acoustic environments. In order to assess its resilience under various room contexts and representative background noise situations, this work combines RIR-based acoustic modeling, NLMS adaptive filtering, and modulation-domain speech augmentation into a single framework.

III. PROPOSED METHODOLOGY

A two-stage acoustic echo cancellation technique is used in the suggested framework. In order to replicate realistic acoustic settings, Room Impulse Responses (RIRs) are first created using the Image Source Method. The microphone signal is created by combining the near-end voice, background noise, and generated acoustic echo. After estimating the dominating acoustic echo using the Normalized Least Mean Square (NLMS) adaptive filter, residual echo and background noise are suppressed while the target speech signal is preserved using modulation-domain speech augmentation.

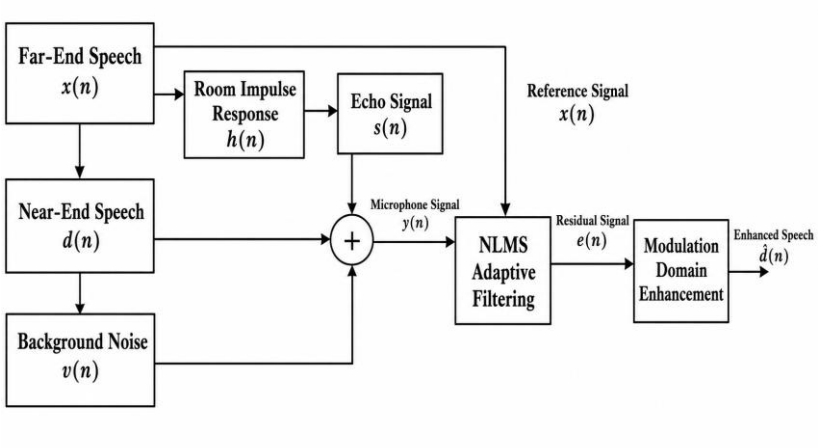


Figure 1: Block diagram of the proposed acoustic echo cancellation framework.

RIRs for acoustic echo modeling are produced using the Image Source Method. To simulate a hands-free conversation situation, the generated echo is mixed with the near-end speech and background noise. Using the far-end speech as the reference signal, the NLMS adaptive filter estimates and suppresses the dominant echo, resulting in the residual signal [2]– [5].

To reduce residual echo and background noise, the residual signal is handled in the modulation domain utilizing STFT, Noise PSD estimation, Wiener gain computation, and ISTFT reconstruction [1], [6]– [10].

Six realistic background noise circumstances, three simulated room environments (RT60 = 0.3 s, 0.6 s, and 0.9 s), and input SNR levels ranging from −10 dB to 20 dB are used to assess the proposed framework. ERLE, Output SNR, SI-SDR, and STOI are used to evaluate performance.

A. Acoustic Echo Generation

The microphone signal consists of the desired near-end speech, acoustic echo, and background noise, and is expressed as

$$y(n) = d(n) + s(n) + v(n) \qquad (1)$$

where $y(n)$is the microphone signal, $d(n)$ is the near-end speech, $s(n)$ is the acoustic echo, and $v(n)$is the background noise.

The acoustic echo is generated by convolving the far-end speech with the room impulse response,

$$s(n) = x(n) * h(n) \qquad (2)$$

where $x(n)$denotes the far-end speech, $h(n)$ is the Room Impulse Response, and * represents the convolution operation. The acoustic environment is modeled using the Image Source Method to generate realistic Room Impulse Responses for echo simulation [2].

B. NLMS Adaptive Filtering

The NLMS adaptive filter estimates the acoustic echo using the far-end speech as the reference signal. The estimated echo is given by

$$\hat{s}(n) = w^T(n)x(n) \qquad (3)$$

Where $\hat{s}(n)$ is the estimated echo, $w(n)$ represents the adaptive filter coefficient vector and $x(n)$is the reference signal vector.
The residual signal is obtained by subtracting the estimated echo from the microphone signal,

$$e(n) = y(n) - \hat{s}(n) \quad (4)$$

Where *e(n)* is the Residual signal after NLMS
The adaptive filter coefficients are updated using the NLMS algorithm as

$$w(n+1) = w(n) + \frac{\mu}{\delta + \| x(n) \|^2} e(n)x(n) \quad (5)$$

where $\mu$is the step size and δ is the regularization parameter. The NLMS algorithm provides stable convergence with low computational complexity, making it suitable for real-time acoustic echo cancellation [3]– [5].

C. Modulation-Domain Residual Echo and Noise Suppression

The residual signal is transformed into the time–frequency domain using the Short-Time Fourier Transform,

$$E(k,m) = \text{STFT}\{e(n)\} \quad (6)$$

The Wiener gain is computed as

$$G(k,m) = \frac{\xi(k,m)}{1 + \xi(k,m)} \quad (7)$$

where $\xi(k,m)$denotes the estimated a priori Signal-to-Noise Ratio.

The gain is estimated from the a priori Signal-to-Noise Ratio and applied to attenuate residual echo and background noise while preserving the desired speech components.

The enhanced spectrum is obtained by

$$\hat{D}(k,m) = G(k.m)\, E(k,m) \quad (8)$$

Finally, the enhanced speech signal is reconstructed using the Inverse Short-Time Fourier Transform,

$$\hat{d}(n) = \text{ISTFT}\{\hat{D}(k,m)\} \quad (9)$$

The modulation-domain enhancement stage combines STFT analysis, Noise PSD estimation, a priori SNR estimation, Wiener gain computation, and ISTFT reconstruction to suppress residual echo and background noise while preserving speech quality [1], [6]–[10].

IV. EXPERIMENTAL SETUP AND PERFORMANCE EVALUATION

A. Experimental Setup

MATLAB R2025b was used to develop the suggested acoustic echo cancellation framework, which was then assessed in simulated acoustic settings created with the Image Source Method [2]. Six realistic background noise circumstances and input SNR levels ranging from −10 dB to 20 dB were used to assess performance in three room contexts (RT60 = 0.3 s, 0.6 s, and 0.9 s). ERLE, Output SNR (SNRout), SI-SDR, and STOI were used for objective evaluation [5], [6].

B. Implementation Parameters

16 kHz far-end and near-end speech signals were used to build the suggested framework in MATLAB R2025b. Room Impulse Responses (RIRs) produced by the Image Source Method were used to imitate acoustic echo [2]. NLMS adaptive filtering [3]– [5] was used to process the microphone signal, and then modulation-domain speech augmentation [1], [6]– [10] was applied.

C. Performance Evaluation Metrics

Echo suppression, speech quality, distortion reduction, and speech intelligibility were assessed using objective metrics such as Echo Return Loss Enhancement (ERLE), Output Signal-to-Noise Ratio (SNRout), Scale-Invariant Signal-to-Distortion Ratio (SI-SDR), and Short-Time Objective Intelligibility (STOI) [5], [6].

1) Echo Return Loss Enhancement (ERLE)

By comparing the microphone signal's power before and after echo suppression, Echo Return Loss Enhancement (ERLE) assesses how well acoustic echo cancellation works. More efficient acoustic echo cancellation is indicated by higher ERLE values [5].

$$ERLE(dB) = 10\log_{10}\left(\frac{E[y^2(n)]}{E[e^2(n)]}\right) \qquad (10)$$

where $y(n)$ is the microphone signal and $e(n)$ is the residual signal after echo cancellation.

2) Output Signal-to-Noise Ratio (SNRout)

After acoustic echo cancellation and residual noise suppression, the output signal-to-noise ratio (SNRout) assesses the quality of the improved speech signal. Better speech enhancement performance is indicated by higher SNRout values [6].

$$SNR_{out} = 10\log_{10}\left(\frac{\sum d^2(n)}{\sum[d(n)-\hat{d}(n)]^2}\right) \qquad (11)$$

where $d(n)$ is the clean (desired near-end) speech signal and $\hat{d}(n)$ is the enhanced speech signal.

3) Scale-Invariant Signal-to-Distortion Ratio (SI-SDR)

The distortion between the clean reference signal and the increased speech signal is measured by the Scale-Invariant Signal-to-Distortion Ratio (SI-SDR), which is insensitive to signal scaling. Better speech reconstruction quality is indicated by higher SI-SDR values [12].

$$SI-SDR = 10\log_{10}\left(\frac{\| \alpha d \|^2}{\| \alpha d - \hat{d} \|^2}\right) \qquad (12) \qquad where \quad \alpha = \frac{\hat{d}^T d}{\| d \|^2}$$

4) Short-Time Objective Intelligibility (STOI)

Following speech augmentation and acoustic echo cancellation, Short-Time Objective Intelligibility (STOI) offers an objective assessment of speech intelligibility. STOI values vary from 0 to 1, with values nearer 1 denoting higher speech intelligibility preservation [11].

V. RESULTS AND DISCUSSION

Three simulated room environments (RT60 = 0.3 s, 0.6 s, and 0.9 s), six representative background noise circumstances, and input SNR values ranging from −10 dB to 20 dB were used to test the suggested framework, which was implemented in MATLAB R2025b. To evaluate its efficacy in voice enhancement and acoustic echo cancellation, both qualitative and quantitative assessments were conducted.

A. Representative Waveform and Spectrogram Analysis

Small Room (RT60 = 0.3 s)

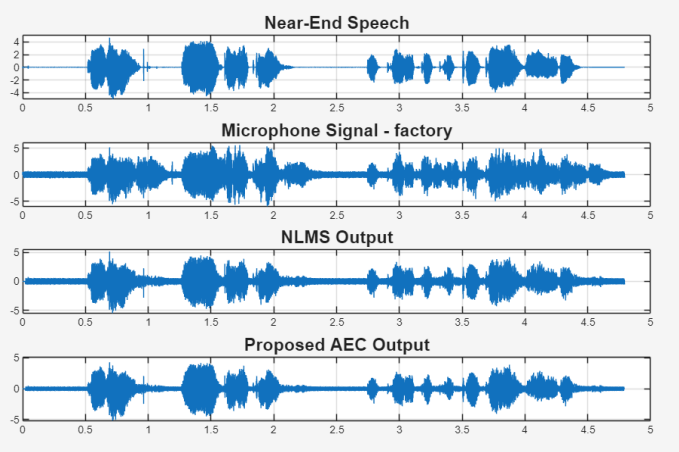


Figure 2. Waveform analysis for the Small Room under Factory Noise at an input SNR of 10 dB.

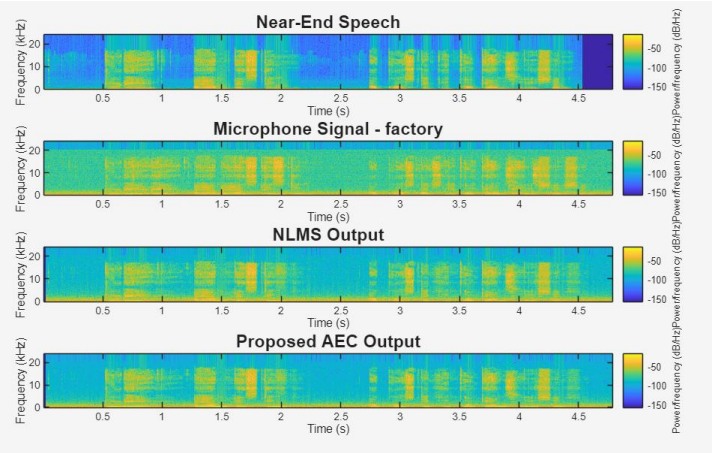


Figure 3. Spectrogram analysis for the Small Room under Factory Noise at an input SNR of 10 dB.

There are fewer reflections and a shorter echo path in the Small Room since it is the least reverberant acoustic environment. With well-preserved speech harmonics in the spectrogram, the suggested framework successfully suppresses the dominant acoustic echo, resulting in a crisper waveform and less residual echo.

Medium Room (RT60 = 0.6 s)

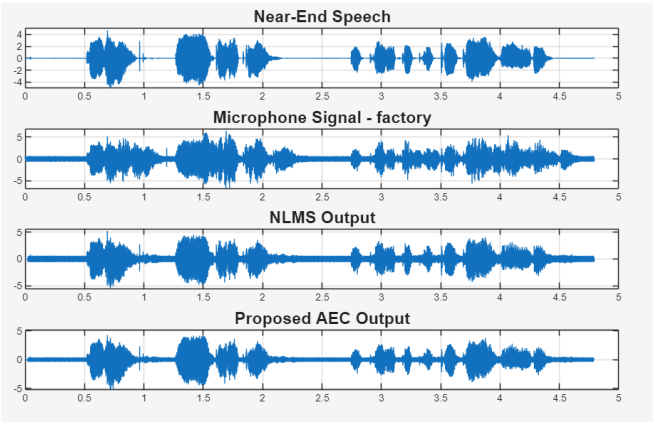


Figure 4. Waveform analysis for the Medium Room under Factory Noise at an input SNR of 10 dB.

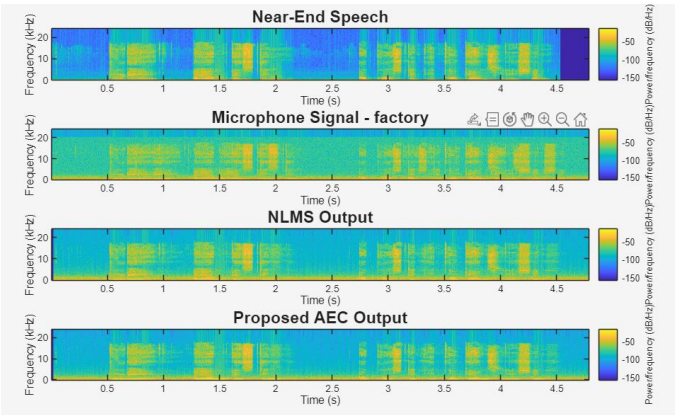


Figure 5. Spectrogram analysis for the Medium Room under Factory Noise at an input SNR of 10 dB.

A more intricate acoustic echo path is produced by the medium room's moderate reverberation. The suggested framework produces a cleaner waveform and better-preserved speech components in the spectrogram by efficiently suppressing acoustic echo and residual interference.

Large Room (RT60 = 0.9 s)

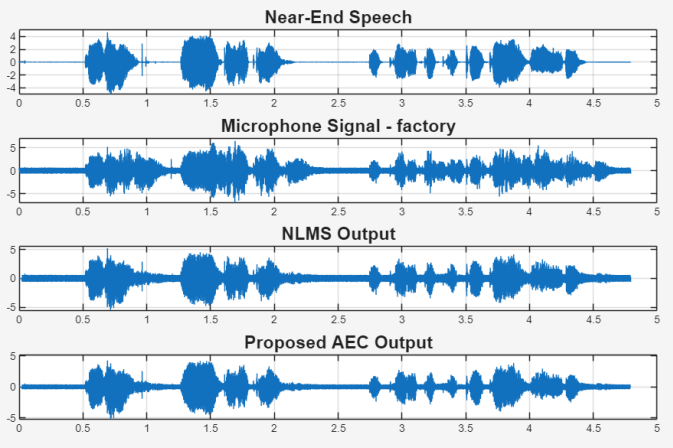


Figure 6. Waveform analysis for the Large Room under Factory Noise at an input SNR of 10 dB.

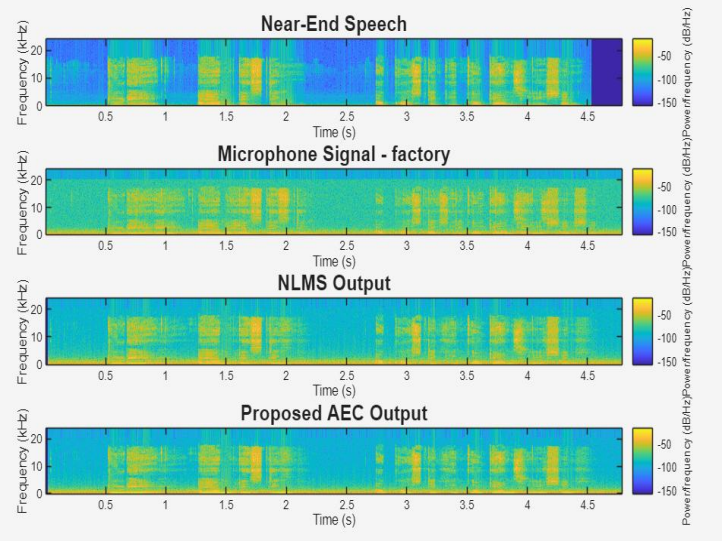


Figure 7. Spectrogram analysis for the Medium Room under Factory Noise at an input SNR of 10 dB.

Because of its longer reverberation period and more reflections, the Large Room is the most difficult acoustic environment. The suggested framework successfully reduces residual interference and acoustic echo, producing a spectrogram with a cleaner waveform and better-preserved voice harmonics.

B. Quantitative Performance Evaluation

The objective performance of the suggested framework under various room environments and background noise situations at an input SNR of 10 dB is summarized in Table I. Efficient suppression of residual echo and background noise is indicated by higher SNRout, SI-SDR, and STOI values. Since background noise largely influences the modulation-domain enhancement stage and ERLE is mostly dependent on the acoustic echo path and the NLMS adaptive filtering step, the virtually constant ERLE values are expected. In general, the suggested framework preserves strong voice enhancement and echo cancellation in all room settings.

Table I Performance of the Proposed Acoustic Echo Cancellation Framework at an Input SNR of 10 dB

Room Environment	Noise Type	ERLE (dB)	SNRout (dB)	SI-SDR (dB)	STOI
Small (RT60 = 0.3 s)	White	13.26	10.67	10.41	0.667
	Babble	13.26	9.39	8.92	0.698
	Factory	13.26	10.53	10.16	0.713
	Cockpit	13.26	10.84	10.57	0.851
	Street	13.26	10.63	10.33	0.542
	Car	13.26	10.76	10.48	0.833
Medium (RT60 = 0.6 s)	White	17.97	10.50	10.22	0.660
	Babble	17.97	8.97	8.43	0.682
	Factory	17.97	10.16	9.76	0.700
	Cockpit	17.97	10.57	10.28	0.861
	Street	17.97	10.36	10.04	0.531
	Car	17.97	10.53	10.23	0.838
Large (RT60 = 0.9 s)	White	16.25	10.16	9.85	0.652
	Babble	16.25	8.71	8.13	0.677
	Factory	16.25	9.92	9.50	0.697
	Cockpit	16.25	10.23	9.92	0.847
	Street	16.25	10.13	9.79	0.525
	Car	16.25	10.17	9.85	0.828

C. Performance Comparison

As the input SNR goes from −10 dB to 20 dB, the performance graphs show that the suggested framework continuously improves speech quality and intelligibility; output SNR, SI-SDR, and STOI increase steadily, and the practically constant ERLE values show reliable acoustic echo cancellation. The robustness of the suggested framework under various reverberation circumstances is further confirmed by the slight variance between the three-room scenarios.

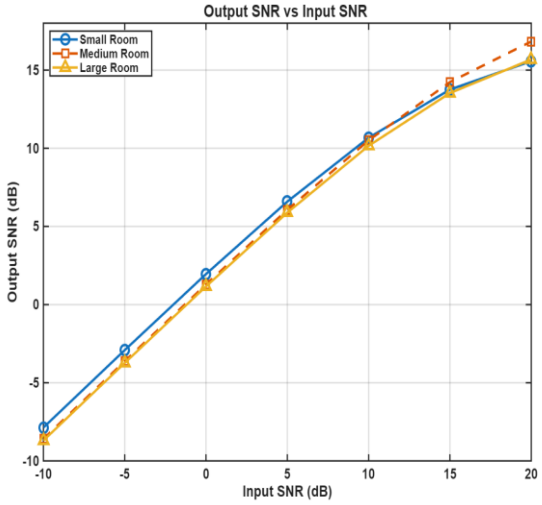


Figure 8. Output SNR versus Input SNR

For the three-room settings, Figure 8 shows how the output SNR changes as the input SNR increases. The output SNR steadily rises from −10 dB to 20 dB, demonstrating successful background noise and acoustic echo suppression. While the Small and Large Rooms function similarly, the Medium Room attains somewhat higher values at greater input SNRs. These findings show that the suggested architecture successfully maintains speech quality in a variety of reverberation scenarios.

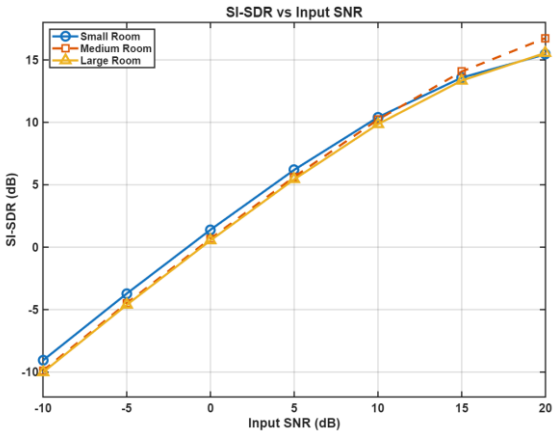


Figure 9. SI-SDR versus Input SNR

The SI-SDR variance for the three-room settings is shown in Figure 9. SI-SDR shows decreased residual echo and voice distortion following processing, increasing consistently with input SNR. The three-room settings show stable speech reconstruction under various reverberation situations, with comparable performance over the assessed SNR range.

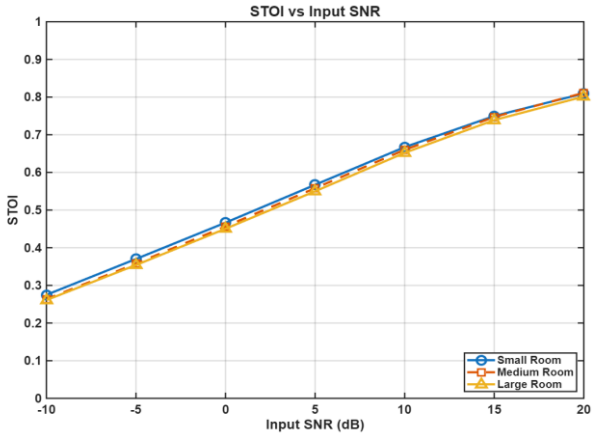


Figure 10. STOI versus Input SNR

For the three-room settings, Figure 10 illustrates how STOI varies as input SNR increases. Following acoustic echo cancellation, speech intelligibility steadily increases from roughly 0.32 at −10 dB to 0.87 at 20 dB, indicating successful preservation of speech quality. The Medium and Large Rooms perform similarly, showing robust functioning under various reverberation circumstances, whereas the Small Room obtains slightly higher STOI values.

D. Discussion

The outcomes shown in Table I and Figures 8–10 show that the suggested framework consistently enhances speech and cancels acoustic echo in a variety of room settings and background noise situations. As input SNR increases, output SNR, SI-SDR, and STOI all gradually improve, demonstrating efficient suppression of background noise and acoustic echo while maintaining speech intelligibility and quality. The resilience of the suggested framework under realistic reverberant settings is further confirmed by the minimal performance variance between the three-room scenarios.

VI. CONCLUSION

For hands-free communication systems, this work introduced a hybrid acoustic echo cancellation architecture that combines modulation-domain voice enhancement, NLMS adaptive filtering, and Room Impulse Response (RIR) simulation. Consistent increases in ERLE, output SNR, SI-SDR, and STOI were shown by experimental evaluation under three room environments, six representative background noise circumstances, and input SNR levels ranging from −10 dB to 20 dB. Effective acoustic echo reduction and enhanced speech quality in various acoustic settings are confirmed by the results.

VII. FUTURE WORK

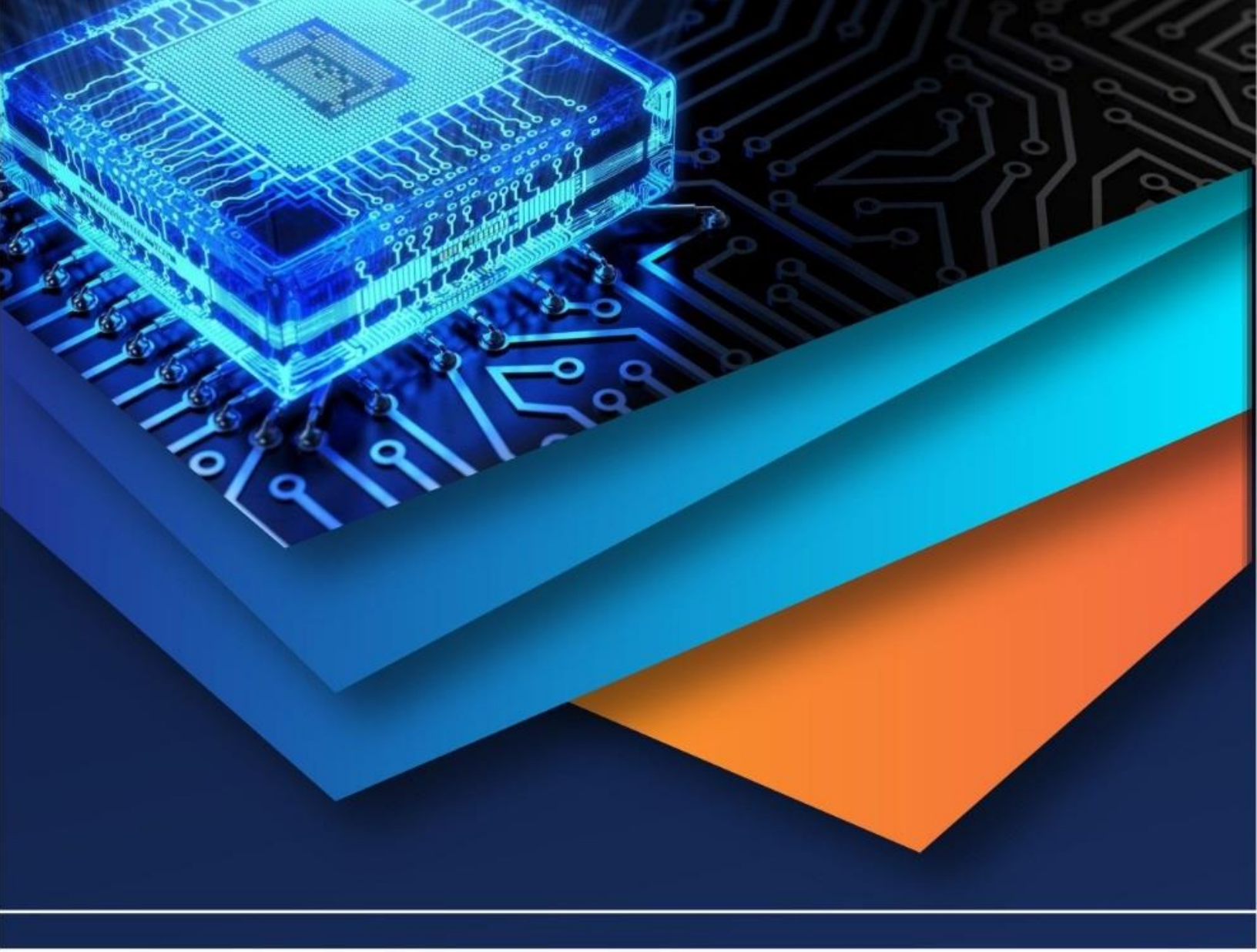
Even if the suggested framework shows good acoustic echo cancellation in simulated acoustic situations, more advancements could be investigated. To enhance performance in dynamic acoustic environments, future research will concentrate on integrating adaptive residual echo suppression and double-talk detection. Furthermore, real-time implementation for useful hands-free communication systems and measured Room Impulse Responses (RIRs) will be used to validate the suggested framework.

VIII. ACKNOWLEDGMENT

The authors would like to thank the Department of Electronics and Communication Engineering of Jawaharlal Nehru Technological University Hyderabad (JNTUH) which provided them the resources and academic help to carry out the study. The authors are heartily thankful to Dr. M. Madhavi Latha for her expert guidance, insightful suggestions and steadfast support while developing this book. Additionally, the authors thank all of the faculty members and colleagues who helped to successfully complete this study, whether directly or indirectly.

REFERENCES

[1] E. P. Jayakumar, P. V. Muhammed Shifas, and P. S. Sathidevi, "Integrated acoustic echo and noise suppression in modulation domain," International Journal of Speech Technology, vol. 19, no. 3, pp. 611–621, Sept. 2016.
[2] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, Apr. 1979.
[3] S. Haykin, *Adaptive Filter Theory*, 5th ed. Upper Saddle River, NJ, USA: Pearson Education, 2014.
[4] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*. Berlin, Germany: Springer, 2001.
[5] P. C. Loizou, *Speech Enhancement: Theory and Practice*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2013.
[6] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.
[7] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 504–512, Jul. 2001.
[8] K. K. Paliwal, K. Wójcicki, and B. Schwerin, "Single-channel speech enhancement using spectral subtraction in the short-time modulation domain," *Speech Communication*, vol. 52, no. 5, pp. 450–475, 2010.
[9] K. K. Paliwal, B. Schwerin, and K. Wójcicki, "Speech enhancement using a minimum mean-square error short-time spectral modulation magnitude estimator," *Speech Communication*, vol. 54, no. 2, pp. 282–305, Feb. 2012.
[10] MathWorks, "Short-Time Fourier Transform (STFT)," MathWorks Documentation.
[11] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, *"A Short-Time Objective Intelligibility Measure for Time-Frequency Weighted Noisy Speech,"* IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2010.
[12] J. Le Roux, S. Wisdom, H. Erdogan, and J. R. Hershey, *"SDR—Half-Baked or Well Done?"* IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY