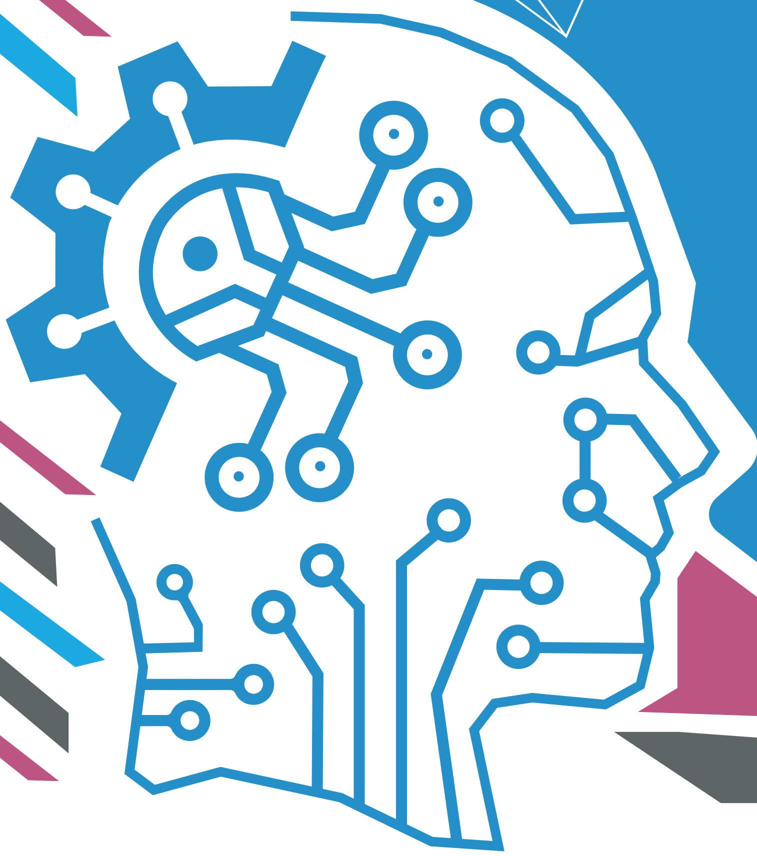
E-BOOK OF EXTENDED ABSTRACT

THE 14TH INTERNATIONAL INVENTION, INNOVATION & DESIGN COMPETITION 2025



14TH IIDES 2025

ENVIRONMENTAL • SOCIAL • GOVERNANCE

UNIVERSITI TEKNOLOGI MARA

RIICAN Research Innovation Industry Community & Alumni Network

E-BOOK OF EXTENDED ABSTRACT

THE 14th INTERNATIONAL
INVENTION, INNOVATION &
DESIGN COMPETITION 2025

Organized by:

Office of Research, Industry,
Community & Alumni Network
UiTM Perak Branch

Perpustakaan Negara Malaysia Cataloguing in Publication Data

No e- ISBN: 978-967-2776-52-9

Cover Design: Dr. Mohd Khairulnizam Ramlie

Typesetting : Georgia

POST-FILTER FOR MULTICHANNEL MALAY SPEECH ENHANCEMENT IN VARIOUS INDOOR NOISY ENVIRONMENTS

Mumtaz Begum Mustafa[1*], Zubair Zaland[1], Miss Laiha Mat Kiah[1]

[1] Faculty of Computer Science and Information Technology, Universiti Malaya, Kuala Lumpur 50603, Malaysia.

mumtaz@um.edu.my

ABSTRACT

Noise suppression is vital in audio processing for improving speech clarity and quality, particularly in indoor environments where reverberations and various noise sources make it difficult to maintain signal clarity. This research examines the performance of the Relative Transfer Matrix (ReTM) using different post-filters in various distances of noise speech sources to tackle the challenges posed by indoor acoustics. ReTM is applied to handle situations with multiple sound sources, capturing complex spatial relationships and improving the noise reduction performance. The experiment used the Malaya-Speech datasets with noise samples from the DEMAND database. The findings indicate notable enhancements in speech quality and intelligibility across various noisy environments and speaker locations, underlining the promise of the post-filters for practical application.

Keywords: Relative Transfer Matrix, Relative-to-maximum Masking, Multichannel Wiener Filter, Temporal Lowpass Filter, Noise Suppression

1. INTRODUCTION

Noise suppression in indoor environments presents a distinct set of challenges due to the acoustic complexities inherent in such spaces. One of the primary issues is reverberation (Wei et al., 2020), which complicates the existing noise conditions. This effect drastically affects technology performance, such as ASR systems and wearable devices for real-life applications. While much of the current research and testing in noise suppression has focused on English-language materials, Nazri et al. (2024), highlighted the need to explore language-specific approaches, suggesting that incorporating Malay-language test materials could yield more contextually relevant improvements for Malaysian users. In response to this gap, the present study aims to examine a new method for enhancing multichannel speech by comparing the Relative-to-Maximum Masking (RMM), Multichannel Wiener Filter (MWF), and Temporal Lowpass Filter (TLF) post-filters with the Relative Transfer Matrix (ReTM) algorithms to tackle the challenges of distant speech sources to improve Malay ASR system performance.

2. METHODOLOGY

2.1 Dataset

This research used the Malaya-Speech dataset (Husein et al., 2020). Malay is a language in the Austronesian (Malayo-Polynesian) language family, spoken as a native language by more than 33,000,000 people distributed over the Malay Peninsula, Sumatra, and Borneo (Nazri et al., 2024).

2.2 Proposed method

Figure 1 shows the ReTM noise suppression algorithm, which is applied to handle situations with multiple sound sources, capturing complex spatial relationships and improving the noise reduction performance. This approach simulates different noisy and reverberant scenarios to assess how well the noise suppression algorithm performs and how much the post-filters improve the performance. ReTM is estimated blindly by analysing the covariance matrix during intervals where speech is absent (Kumar et al., 2024), disentangling desired speech signals from unwanted background noise (Manamperi et al., 2024). RMM takes this further by implementing a masking technique that minimises residual noise (Lin et al., 2020). MWF (Doclo et al., 2007) significantly improves the quality of the desired signal, cutting through the surrounding noise. MWF maintains performance in changing environments, resulting in a higher SNR output. On the other hand, TLF (Liu et al., 2009) employs a simple yet effective moving-average filter, designed to highlight the lower modulation frequencies inherent in speech signals.

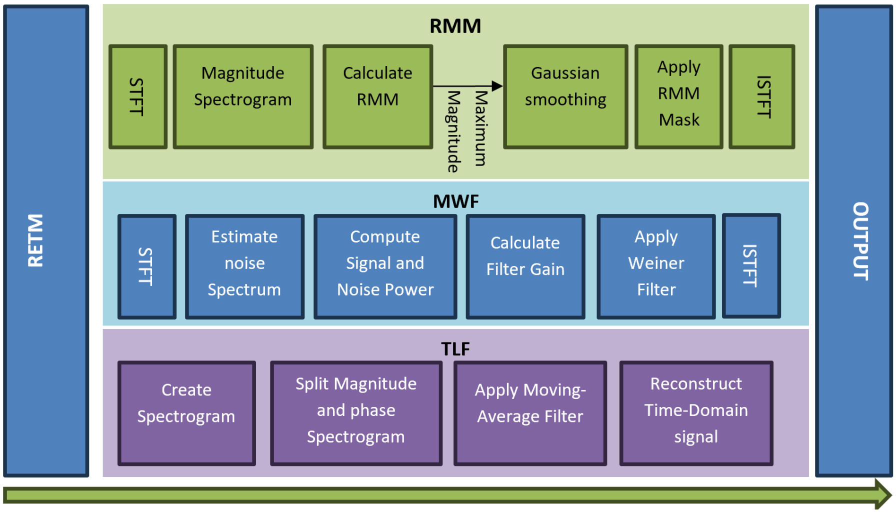


Figure 1 ReTM with RMM, MWF, and TLF post-filters.

2.3 Experimental design

ReTM utilises timestamps of the non-speech segments in the input signals obtained using voice activity detection, using a confidence threshold of 0.8 and a regularisation value of $1e-5$ for calculating the Relative Transfer function. The RMM was configured with a window size of 128ms and 85% overlap between consecutive windows, and the noise floor mask was set at 0.2. The TLF was set up using 256 samples per segment, with 78% overlapping samples between segments. Finally, MWF was configured using a window size of 2048 for STFT and an overlap of 75%. The Noise source was placed at a constant distance of 1 meter and 45 degrees from the microphones, while the speech source was recorded at arbitrary distances of 1.2, 2.5, 3.8, and 5.1 meters from the microphones, as shown in Figure 2.

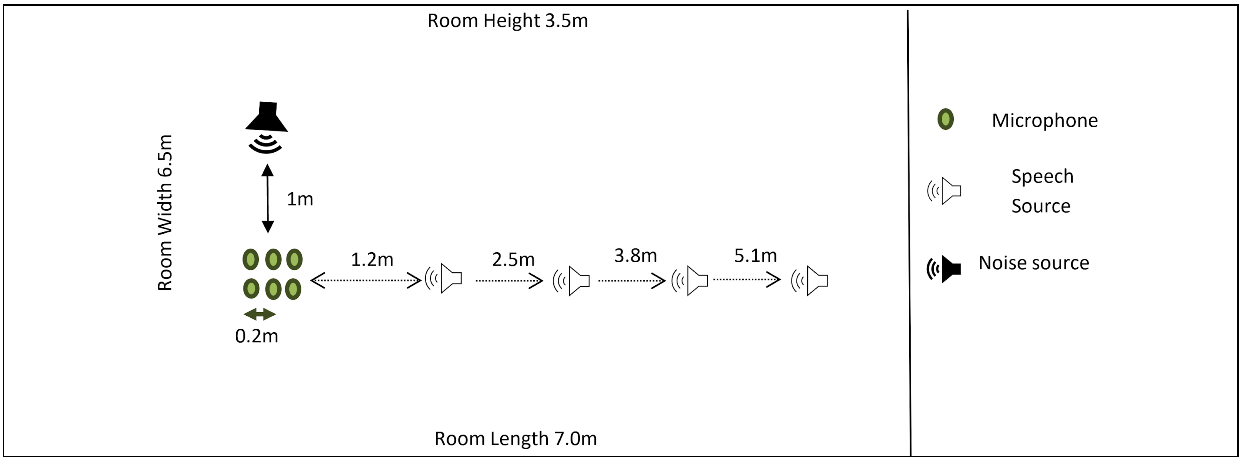


Figure 2 Configuration of input, speech, and noise sources in the experimental recordings

Table 1 summarises the experimental design in this research. The Speech enhancement metrics used in these experiments are the Perceptual Evaluation of Speech Quality (PESQ) and Short-time Objective Intelligibility (STOI). These performance scores were recorded in terms of percentage of improvements in PESQi and STOIi scores compared to their input scores.

Table 1 Summary of the experimental design.

Post-filter	Window Size (ms)	Overlap (%)	Source Distance (m)	Test Dataset	Noise Dataset	Noise Type

RMM	128	85	1.2, 2.5, 3.8, 5.1	Malaya-Speech	DEMAND	Cafe and Office
TLF	256	78				
MWF	2048	75				

3. FINDINGS

Table 2 presents the results of the tests conducted on the Malaya-Speech datasets under two noise conditions: Office and Café, comparing the baseline (ReTM with no post-filter) with the proposed post-filter. The RMM has the highest improvement across all distances, ranging between 10.79% and 4.32% for office noise and 11.18% and 4.78% for Café noise. The TLF post-filter yielded the best STOI improvement scores from 9.35% to 3.32% for the office noise. The STOIi scores for café noise showed very low improvement overall, with the RMM scoring highest at 1.2 meters, obtaining 1.73% improvement. At distances beyond 1.2m, the MWF scored higher improvement, ranging between 1.15% and 0.61%. At distances between 2.5 and 5.1 meters, MWF again rose to the top, scoring 1.23% to 0.65% improvement, respectively.

Table 2 Test Results for Office and Café noises for ReTM (-) with Post-filters.

Dataset	Noise Type	Post-Filter	PESQi				STOIi			
			1.2	2.5	3.8	5.1	1.2	2.5	3.8	5.1
Malaya-Speech	Office	-	**5.32**	**2.81**	**1.44**	**3.08**	**8.24**	**5.54**	**4.14**	**2.45**
		RMM	10.79	3.81	3.27	4.32	8.60	5.07	3.19	0.97
		TLF	5.68	3.21	1.97	3.46	9.35	6.82	5.29	3.32
		MWF	5.33	2.98	1.82	3.16	8.12	5.81	4.44	2.91
	Café	-	**5.61**	**3.39**	**1.83**	**3.21**	**1.66**	**1.09**	**0.84**	**0.51**
		RMM	11.18	4.64	4.04	4.78	1.73	0.99	0.64	0.20
		TLF	5.61	3.39	1.83	3.21	1.66	1.09	0.84	0.51
		MWF	5.85	3.60	2.25	3.39	1.66	1.15	0.91	0.61

4. CONCLUSION

The overall findings indicate that no single post-filter performs optimally in every situation. RMM provides the most significant improvements in perceptual quality at close distances but has significant challenges with intelligibility at longer ranges and requires the most processing power in real-time. TLF consistently produces best intelligibility in short to mid-range scenarios (1.2–3.8 meters) and manages a favorable balance between enhancement effectiveness and computational efficiency, making it suitable for real-time use. Although MWF is resource-intensive, it is particularly effective in preserving intelligibility over longer distances, especially in the constantly changing characteristics of café noise. Therefore, the choice of post-filter should be guided by the specific context of the application, including the intended distance, desired enhancement type, and hardware resource constraints.

ACKNOWLEDGEMENT

The Ministry of Higher Education financially supported this research under the Fundamental Research Grant Scheme (FRGS/1/2023/ICT09/UM/02/1).

REFERENCES

Doclo, S., Spriet, A., Wouters, J., Moonen, M. (2007). Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction. *Speech communication 49*(7-8),636–656. https://doi.org/10.1016/j.specom.2007.02.001

Husein, Z. (2020). *Malaya-Speech, Speech-Toolkit library for Malaysian language, powered by PyTorch* [Computer software]. GitHub. https://github.com/mesolitica/malaya-speech

Kumar, M., Birnie, L., Abhayapala, T., Holzinger, S.A., Bastine, A., Grixti-Cheng, D., Samarasinghe, P. (2024). Speech denoising in multi-noise source environments using multiple microphone devices via relative transfer matrix. *2024 European Signal Processing Conference (EUSIPCO)*, 281–285.

Lin, C-L., Lin, Z-Q., Wang, S-S., Tsao, Y., Hung, J-W. (2020). Exponentiated magnitude spectrogram-based relative-to-maximum masking for speech enhancement in adverse environments. In *2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)* (pp. 1-2). IEEE. DOI: 10.1109/icce-taiwan49838.2020.9258262.

Liu, K-Y., Wang, S-S., Tsao, Y., Hung, J. (2019). Speech enhancement based on the integration of fully convolutional network, temporal lowpass filtering and spectrogram masking. In C.-Y. Chiag, M.-Y. Day, & J.-T. Chien (Eds.), *Proceedings of the 31st Conference on Computational Linguistics and Speech Processing (ROCLING 2019)* (pp. 226–240). The Association for Computational Linguistics and Chinese Language Processing (ACLCLP). https://aclanthology.org/2019.rocling-1.22/

Manamperi, W.N., Abhayapala, T.D. (2024). Relative transfer matrix for drone audition applications: Source enhancement. In *APSIPA ASC 2024 - Asia Pacific Signal and Information Processing Association Annual Summit and Conference 2024* (APSIPA ASC 2024 - Asia Pacific Signal and Information Processing Association Annual Summit and Conference 2024). Institute of Electrical and Electronics Engineers Inc.. https://doi.org/10.1109/APSIPAASC63619.2025.10848886

Nazri, M. Z., Mohamad Zaini, N. S., Kamaruddin, N. N., Rameezal, R.T.N., Musa, N.F., Basar, N., & Razal, M.R. (2024). Synthesis, characterizations and multivariate data analysis of non-and nitrile-functionalized N-heterocyclic carbene complexes of silver (I) and palladium (II). *Scientific Research Journal, 21*(2), 89-117.

Wei, Y., Zhang, K., Wu, D., Hu, Z. (2020). Comparison of conventional enhancement and separation methods in indoor environments. In *Proceedings of the 2020 9th International Conference on Computing and Pattern Recognition (ICCPR ’20)* (pp. 383–389). Association for Computing Machinery. https://doi.org/10.1145/3436369.3436488

E-Book of Extended Abstract THE 14th INTERNATIONAL INVENTION, INNOVATION & DESIGN COMPETITION 2025

e ISBN 978-967-2776-52-9

Unit Penerbitan UiTM Perak
.

(online)