
Research article

Enhancing Speaker Verification Performance in Noisy Environments via Parallel Model Combination

Duraïd Y. Mohammed ^{1,*}, Ahmed Aljuboori ²

¹ School of education for women, Al-Iraqia University, Baghdad, Iraq

² College of Education for Pure Science Ibn-Al-Haitham, University of Baghdad, Baghdad, Iraq

* **Correspondence:** duraidyehya19@gmail.com

Keywords:

speaker verification, HMM, robustness, PMC, noise

Important Dates:

Received: 10 June 2026

Revised: 30 June 2026

Accepted: 7 July 2026

Online: 15 July 2026

Copyright © 2026 by the authors. Published under Creative Commons Attribution (CC BY) license.

ABSTRACT

In real-world speaker applications, background and additive and convoluted noise and reverberation cause a mismatch between the training and recognition environments, resulting in poor performance. Parallel Model Combination (PMC) has proven to be an effective method for mitigating the effects of additive noise. Hidden Markov Model (HMM)-based speaker recognizers have been successfully improved using Parallel Model Combination (PMC). The results of using PMC to compensate for additive noise in HMM-based text-independent speaker verification are presented in this paper. Speech data were obtained from the SALU-AC databases. The findings of using PMC to correct for additive noise in HMM-based text-independent speaker verification are presented in this article. Equal Error Rates (EER) for noise-corrupted speech are given for various signal-to-noise ratios (SNRs) and noise sources. For example, at 10dB SNR, the average EER for speech in an operations room noise decreased from 6.7% uncompensated to less than 4.6% using PMC. Finally, it is shown that the precise value of the parameter that controls the relative amplitudes of the PMC model's speech and noise components does not affect speaker verification accuracy. Speaker verification experiments in artificial conditions show the efficiency of using Parallel Model Combination (PMC) in reducing equivalent EER error rate and DET trade-off error detection.

1. Introduction

When there is a mismatch between the training and testing environments, automated speaker recognition (ASR) performance quickly decreases. It has been shown that ASR systems perform extremely poorly when evaluated in an acoustic environment that differs from the one they were trained in. As a result, the need for robust speaker recognition systems is growing as they are used in real applications. Applications such as voice recognition over the phone, in automobiles, on a manufacturing floor, or outdoors require high environmental resilience[1, 2]. After several years of work, present speaker recognition systems have realized an acceptable improvement, assuming that the training and retrieval speech signal length is adequate in addition to a significant signal-to-noise ratio (SNR) and reverberation time[3-5]. Several efforts have been made to address these challenges, to name a few. AL-KARAWI's study has proved that training on the fly can reduce the harmful effect of reverberation on verification performance [6, 7], feature extraction space, and i-vector[8-10]. The author in [11-13] used a multi-condition training technique to decrease the impacts of noise and reverberation on verification performance. The GMM-UBM framework characterizes a conventional speaker recognition method [14-17]. Noise-corrupted speech

signals are widely known to cause higher speech and speaker verification mistakes due to the mismatch between training and test conditions and information loss. As a result, significant work has been spent developing effective noise correction methods. These methods are divided into speech pre-processing and recognition stage modification. The first class of techniques attempts to pre-process damaged speech to provide parameters typical of clean speech[8, 18].

Spectral subtraction and mapping [19, 20] are examples of designs in this area. HMM decomposition [21] and Parallel Model Combination (PMC) [22-24] are examples of methods that alter the recognition step. Methods based on pre-processing are typically computationally more accessible. However, they have the drawback of not retrieving any critical information lost during the pre-processing step for use in recognition. The work presented in this article focuses on the second method, namely Parallel Model Combination (PMC), which compiles speech and noise HMMs into a single composite HMM before recognition[25, 26]. PMC has been effectively used in HMM-based automated speech recognition, which has been demonstrated to enhance recognition performance on speech with additive noise[23, 27, 28]. The findings of studies that explore the usefulness of PMC for noise-robust HMM-based text-dependent speaker verification are presented in this article. As test data, clean speaker verification data from the SALU-AC) was mixed with bubble noises at various signal-to-noise ratios[29]. Using PMC, this data was processed using clean speech models coupled with suitable noise models at the proper mixing levels[30, 31]. The findings indicate that PMC reduces equal error rates significantly. Speaker verification aims to minimize the consequences of such a mismatch to achieve recognition accuracy near that achieved under the matched condition as feasible. A wide range of solutions to the issue of robust speaker verification have been suggested. The Parallel Model Combination (PMC) method will focus on this article since it is a practical approach for minimizing the impacts of additive noise. The paper is organized as follows. Section 2: Rationale and System Architecture, Section 3: Experiment Setup, Section 4: Experiment Results, and finally, conclusions are in Section 5.

2. Rationale and System Architecture

Speaker verification aims to validate a subject's stated identification by using individual variations in their speech. It is essential to differentiate between text-dependent speaker verification, which uses speech matching to recognize text, and text-independent speaker verification, which uses unconstrained vocabulary [6, 7, 17]. The current research is focused on the former. Text-dependent verification is a more straightforward issue to solve, and it is susceptible to word-level acoustic modeling methods from automated speech recognition, particularly the use of HMMs. After deciding to utilize HMM methods, it is reasonable to wonder if HMM noise compensation techniques from voice recognition can also be effectively used. PMC[32] is the noise-compensating method under investigation in this research.

2.1 Speaker recognition

A speaker recognition system is designed to identify and verify an individual's identity based on their unique vocal characteristics. This biometric security system analyzes and captures distinct features of a person's voice, such as pitch, tone, rhythm, and pronunciation[7, 11]. Speaker recognition systems can be categorized into two main types: verification and identification. In the verification mode, the system confirms whether the provided voice sample matches the pre-enrolled speaker's reference template[33, 34]. On the other hand, in the identification mode, the system determines the speaker's identity from a database of known speakers. This technology finds applications in various fields, including security access control, telecommunications, and forensic investigations, where identifying individuals based on their voice patterns is crucial. Advances in machine learning and signal processing techniques have significantly improved the accuracy and reliability of speaker recognition systems, making them integral in modern security and authentication systems. Figure 1 shows the framework of the speaker recognition system.

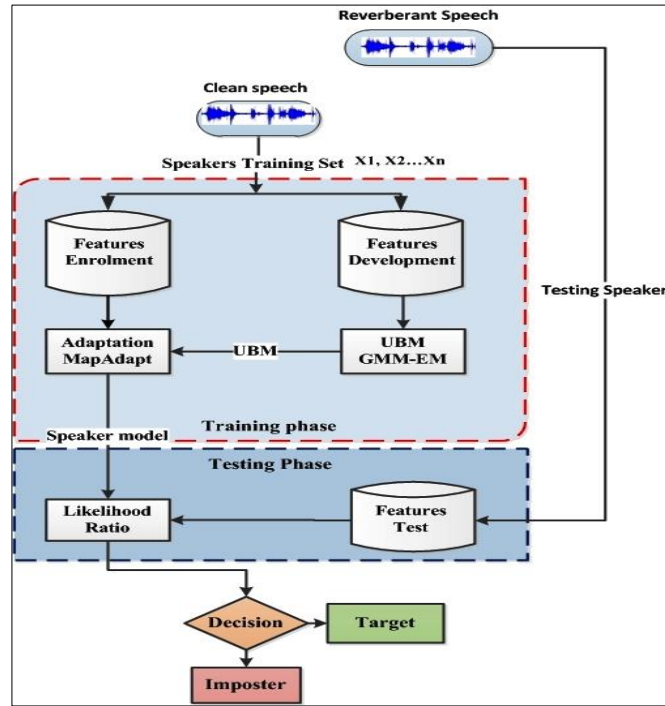


Figure 1: The framework of the speaker recognition system

2.2 Feature Extraction

Mel Frequency Cepstral Coefficients (MFCC) feature extraction is widely used in speech and audio signal processing. It involves capturing essential aspects of the human auditory system's response to different frequencies, making it an effective method for representing audio signals. The process begins with dividing the audio signal into short overlapping frames. The power spectrum is computed for each frame, followed by applying a mel filterbank to emphasize specific frequency ranges[5, 35, 36]. The logarithm of the filterbank energies is then taken, and the discrete cosine transform (DCT) is used to obtain a set of cepstral coefficients. The resulting MFCCs represent the spectral characteristics of the audio signal and are particularly useful for speech and speaker recognition tasks. They compactly encode the signal's frequency content information, enabling effective feature representation for subsequent machine learning or pattern recognition algorithms[11, 30]. MFCCs have become a standard in audio processing applications, contributing significantly to the success of systems that analyze and recognize speech and other audio signals.

Mel Frequency Cepstral Coefficients (MFCC) calculation is crucial in extracting distinctive features from audio signals, particularly speech and speaker recognition[13, 16, 27]. The process involves several key steps to capture the essential characteristics of the human auditory system. Firstly, the audio signal is divided into short frames, typically around 20 to 40 milliseconds long, with each frame overlapping the previous one. Subsequently, a Fourier transform is applied to convert the signal from the time domain to the frequency domain. A Mel filter bank emphasizes specific frequency bands, mimicking the non-linear human ear response to different frequencies[6]. The logarithm of the filterbank energies is taken, and then the Discrete Cosine Transform (DCT) is applied to decorrelate the resulting coefficients. The coefficients obtained through this process are the MFCCs, which represent the spectral features of the audio signal in a compressed form. The final set of MFCCs is used as a feature vector for various machine learning and pattern recognition tasks, playing a fundamental role in applications such as speech recognition, speaker verification, and audio signal processing[37].

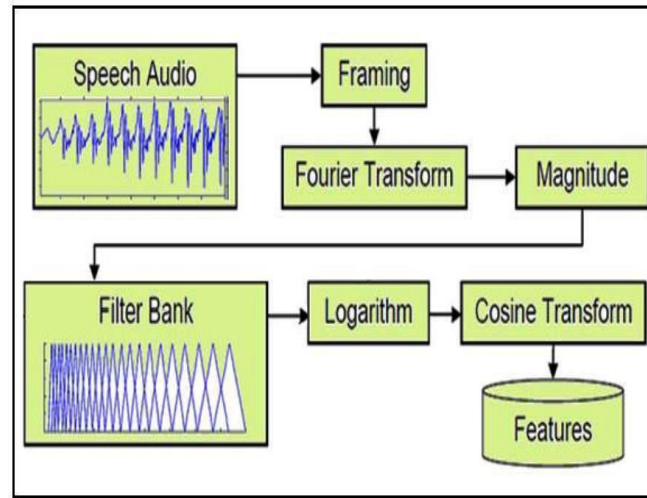


Figure 2 MFCC extraction procedure

2.3 Parallel model combination (PMC)

PMC is founded on the assumption that noise compensation should occur at the pattern processing stage of voice or speaker recognition rather than during parameterization. The judgment that a component of the input is 'noise,' in particular, should come from the recognition process rather than precede it. In this manner, all the information in the voice signal is preserved and used for proper verification. Despite being comparable to HMM decomposition[21, 38], PMC has the advantage of working in the Cepstral domain and, therefore, inheriting the benefits of parameter decorrelation. Let σS and σN be single Gaussian states of a 'clean' speech HMM and a noisy HMM in the cepstral domain, respectively. Assume that the means and variances of σS and σN are represented by

$$\{\mu_S^C, \sum_S^C\} \text{ and } \{\mu_N^C, \sum_N^C\} \quad (1)$$

PMC generates a combined cepstral state $\sigma S \otimes \sigma N$ by inverse-transforming

$$\{\mu_S^C, \sum_S^C\} \text{ and } \{\mu_N^C, \sum_N^C\} \quad (2)$$

Into the linear spectral domain (through the log spectral domain), properly merging them into a single distribution, and then projecting this distribution back into the Cepstral domain. PMC makes many assumptions throughout this procedure. It is assumed that speech and noise are independent and additive and the two lognormal distributions are log-normal throughout the combining process. PMC can adjust for static and dynamic (velocity and acceleration) factors usually included in a Cepstrum-based representation for speech recognition. However, only static characteristics were modified in this research since it was thought that correcting for the delta and acceleration parameters would only result in minor increases in performance because the differencing procedure would eliminate many of the speaker-dependent features. In the log spectral domain, the speech means and covariance matrix are given by:

$$\mu_S^l = C^{-1} \mu_S^C \quad (3)$$

$$\sum_S^l = C^{-1} \sum_S^C (C^{-1})^T \quad (4)$$

Where C is the cosine transform defined by the matrix

$$C_{ij} = \cos\left(\frac{i(j - 0.5)\pi}{B}\right) \quad (5)$$

Similar expressions exist for the parameters of the noise state. The mean and covariance matrices for the state $\sigma S \otimes \sigma N$

σN are then given by:

$$\mu_{S \otimes N}^l = \log(\exp(\mu_S^l) + g \exp(\mu_N^l)) \quad (6)$$

$$\mu_{S \otimes N}^l = \log(\exp(\sum_S^l) + g \exp(\sum_N^l)) \quad (7)$$

Again, g is the matching term that determines the signal-to-noise ratio.

2.4 The log-add PMC technique

PMC is a noise compensation scheme based on models for robust speech recognition. PMC aims to change the parameters of a set of HMM-based acoustic models estimated in a clean environment to represent the speech spoken in the current operating environment accurately. The technique assumes that a clean speech model is available and that a simple additive noise model can be estimated online to characterize the actual noise conditions. The first step in any robust recognition technique is defining a distortion or mismatch function that describes the effects of noise on clean speech. According to the most basic and widely used model, the original speech signal can be distorted by both additive and convolutional noise. Thus, the fundamental assumption underlying PMC is that speech and ambient noise are additive in the linear spectral domain, while speech and channel distortion are multiplicative.

Furthermore, it is assumed that the convolutional noise remains constant over time. Another necessary assumption in the PMC framework is that the frame/state alignment, which is used to generate the speech models from the clean speech data, is unaffected by the addition of noise. As a result, the mismatch function in the log spectral domain can be represented as

$$\theta_1^l(t) = \log(g \exp(S_1^l(t)) + \exp(N_1^l(t))) \quad (8)$$

We only consider the Log-Add Approximation among the various PMC approximations, and it is the most straightforward and computationally efficient PMC implementation. The model variances are assumed to be negligible, so their effects on the estimates are ignored. This allows us to obtain a simplified and appropriate mismatch function for the static and dynamic coefficients that only adapts the HMM means while leaving the variances unchanged from the clean models [39]. It is important to note here that a simple noise HMM should be used to reduce both the computational overhead of the PMC technique and the recognition time.

2.5 Noise Model Selection

As previously mentioned, a noise model estimate is required to describe the noise circumstances of the present use environment to apply the PMC methods. Before the model adjustment procedure, the noise HMM is learned during the recognition step. This necessitates sufficient ambient noise data to generate an accurate noise HMM estimate. Non-speech frames inside the utterance may be used to acquire the required ambient noise data (speech pauses). As a result, to conduct noise data selection in a practical recognition system, an appropriate method, such as a dependable Voice Activity Detector (VAD), is needed to distinguish speech from stationary and non-stationary sounds in the operating environment. Assuming that VAD has high speech classification accuracy, another significant element in the effectiveness of noise compensation methods is the quantity of noise data provided. The duration of noise data in real-time recognition systems is a crucial characteristic that should be kept as short as feasible. Usually, the noise model is determined using noise data of finite length. A segment that is too short cannot be utilized since the model estimate would also be unreliable. The methods employed in compensation strategies to estimate the additive noise model acquire the test noise samples by capturing the required background noise before the user begins to speak. The proposed environmental compensation method aims to achieve high identification accuracy using a limited quantity of environment-specific noise data and a low computing cost. The suggested process eliminates the online noise HMM training required for PMC scheme implementation. The fundamental aim is to create a noise model selection that may be utilized in real-time robust speaker recognition. The primary

operations are as follows[40]:

- A. A single-state multiple-mixture general noise HMM model is trained offline using a noise database. This noise database must be representative of our possible operational scenarios.
- B. At the recognition phase, we use the N first frames of the utterance to select one or more Gaussian mixtures as the actual noise model for the compensation process. So far, for this selection, we have tried two different approaches.

The noise selection method described above is shown in Figure 1.

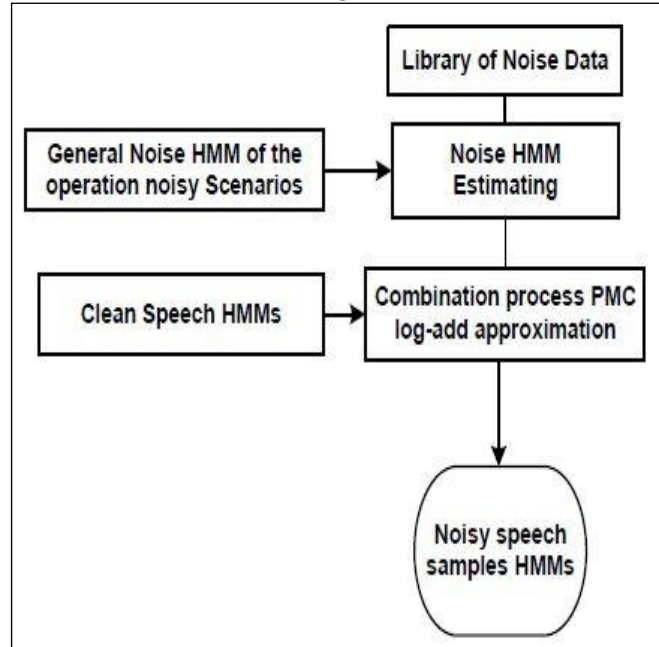


Figure2 The proposed method for noise model selection

In generating the noise model for the generic Hidden Markov Model (HMM), a set of Gaussian distributions, ranging from one to the number of components in the generic noise HMM, is selected using the specified approach. The generation of the essential noise HMM necessitates the merging or mixing of these diverse Gaussian distributions. Various methods can be employed for this purpose, and in this context, a straightforward approach was adopted – a simple average across all the selected distributions. This method computes the mean of the noise HMM, contributing to creating a representative and comprehensive noise model. This approach ensures that the noise HMM encapsulates the essential characteristics derived from the diverse Gaussian distributions, providing a robust representation of the underlying noise components in the system.

$$\hat{\mu} = 1/M \sum_{m=1}^M \mu_m \quad (9)$$

where M is the number of selected Gaussians.

3. Experiment Setup

3.1 Speech dataset

In this experimental setup, a cohort of eighty participants, evenly distributed with 40 men and 40 women, was selected from the SALU-AC databases [41]. The clean voice samples were meticulously recorded in a fully anechoic home environment within a designated chamber at the University of Salford, United Kingdom. This chamber, measuring 78.408m³, maintained a background noise level of approximately -12.4 dB and had a cut-off frequency of 100Hz. Each participant contributed ten utterances, each lasting around 20 seconds. To facilitate training, the recordings were segmented into 5-second intervals. Eight samples (40 seconds) were selected for enrolment purposes, and two utterances from distinct records for each speaker were utilized to construct a test database. The primary objective of this study was to investigate talker recognition concerning the environment, emphasizing the use of brief utterances. Vocal activity detection techniques were employed to eliminate salience

components and enhance accuracy. The voice signal's sample rate was standardized to 16 kHz, ensuring consistency and reliability across all recordings.

3.2 Evaluation methods

In statistical testing assumptions, two types of errors play a crucial role in evaluating speaker recognition performance and other biometric frameworks. One of the primary methods employed for this assessment is the Equal Error Rate (EER). In speaker verification, the ultimate goal is to minimize both types of mistakes: the false rejection rate (FRR) and the false acceptance rate (FAR). Achieving optimal system performance entails ensuring that both errors are equal. The Equal Error Rate (EER) represents the percentage value of FRR and FAR when these two errors are identical. This rate can be calculated graphically by monitoring the false acceptance percentage or the percentage of false rejection at the point of intersection of a line and the Detection Error Tradeoff (DET) curve. The EER serves as a critical metric for determining the quality of a speaker recognition system, where a lower EER value indicates superior performance. Thus, minimizing both false rejection and false acceptance rates contributes to the effectiveness of the speaker recognition system [17]. The DET shows the FAR and FRR percentages based on the acceptance threshold setting. As a result, decreasing the FRR likelihood increases the FAR probability and vice versa.

3.3 Mixing Procedure

This study enriched the database above by incorporating diverse noise levels and introducing noisy utterances characterized by signal-to-noise ratio (SNR) values ranging from 20 to -5 dB. The augmentation process involved the development of a specialized audio mixer, meticulously crafted using MATLAB code. This innovative mixer seamlessly integrated anechoic speech, adjusting the signal strength to emulate real-world scenarios featuring various background noise levels. The mixer created a comprehensive test environment encompassing a spectrum of SNR values, facilitating a thorough evaluation of the system's robustness and performance across various challenging acoustic conditions. This deliberate approach ensures that the speaker recognition system undergoes scrutiny in conditions that closely mimic the intricacies of everyday settings, providing crucial insights into its practical applicability and overall reliability [4]. Following this, the speech signal underwent a meticulous integration process with the appropriately scaled noise signal, creating noise-polluted speech signals characterized by signal-to-noise ratio (SNR) intervals of 6 dB. The scaling factor required to achieve a specific SNR varied based on individual speakers and voice files. An average scaling factor, denoted as "g," was calculated for all speakers at each SNR to maintain consistency and accuracy. Subsequently, this calculated scaling factor was used to construct the Parallel Model Combination (PMC) models corresponding to each SNR. The mixing process involved several essential steps, starting with addressing normalization concerns to ensure the accurate addition of the requisite number of speech and noise samples, thereby preventing misinterpretation. The perceived levels of the mixed signals were equalized through standard methods, ensuring comparability of composite samples by processing them to have equal Root Mean Square (RMS) levels. This meticulous procedure addressed normalization concerns and facilitated the creation of a realistic and standardized set of noise-polluted speech signals. This approach aimed to provide a comprehensive and nuanced evaluation of the speaker recognition system's performance under diverse acoustic conditions by maintaining consistent RMS levels across the generated signals. Such a standardized set allows for a more thorough understanding of the system's robustness and efficacy across varying background noise levels.

In the subsequent step, a comprehensive dataset was assembled, encompassing 800 utterance samples derived from 80 speakers (with 10 utterances per speaker, evenly distributed between 40 male and 40 female participants). The novel addition introduced babble noise at varying signal-to-noise ratio (SNR) levels, spanning from -5 dB to 20 dB. These diverse and noisy utterances served as the validation set for evaluating the efficacy of the proposed

speaker recognition technique. Leveraging the recorded babble noise, with a sampling rate of 44.1 kHz, the speech signal underwent a meticulous integration process with appropriately scaled noise signals. This meticulous procedure ensured the generation of noise-polluted speech signals characterized by SNR intervals of 6 dB, with the scaling factor tailored to each speaker and voice file. This approach contributes to a robust evaluation, capturing the system's performance nuances under different acoustic conditions, thereby providing valuable insights into the technique's versatility and effectiveness.

3.4 HMM System

The Hidden Markov Model Toolkit (HTK) framework was utilized for HMM training and recognition and was readily adapted to incorporate PMC. Using a 25ms Hamming window, the speech was parameterized into a 39-dimensional representation based on Mel-frequency Cepstral Coefficients (MFCC). This included the 0th-order energy term plus each coefficient's velocity (Δ) and acceleration (Δ^2).

4 Experimental Results

In evaluating the proposed noisy model selection techniques, our initial baseline system yielded an Equal Error Rate (EER) of 1.2% when applied to clean verification speech. The objective of subsequent experiments was to approach this baseline EER value as closely as possible under various signal-to-noise ratios (SNRs). The experiments were designed to systematically test the impact of additive noise on the system's performance. The baseline system's EER with clean verification speech served as a benchmark at 1.6%. This baseline performance, outlined in Table 1, established a reference point for assessing the effects of noise on the system's accuracy and reliability. The ensuing experiments, conducted with clean and noisy test files, aimed to gauge the efficacy of the proposed noisy model selection techniques and their ability to mitigate the impact of additive noise on the verification system. The reported findings provide valuable insights into the system's adaptability and performance across varying acoustic conditions[40].

Table 1 shows the baseline performance.

Speech length	EER %						
	clean Speech	Noisy speech (SNR(dB))					
		20%	15%	10%	5%	0%	-5%
5s	1.6	2.3	3.4	6.7	11.4	14.2	17.6

The second experiment was designed to investigate the degradation in performance when utilizing noise-corrupted speech for verification without the application of Parallel Model Combination (PMC). In this scenario, the goal was to understand how noise impacts the verification system's effectiveness by subjecting the speech data to various noise levels without the compensatory mechanism provided by PMC. The experiment aimed to highlight the vulnerabilities and limitations of the system when confronted with real-world acoustic challenges. This analysis becomes crucial in assessing the system's robustness and reliability in scenarios where noise is an inevitable factor, and the outcomes contribute valuable insights into the system's performance under adverse acoustic conditions.

Table 2 shows the system performance using PCM

Speech length	EER %						
	clean Speech	Noisy speech (SNR(dB))					
		20%	15%	10%	5%	0%	-5%
5s	1.6	2.0	2.9	4.6	10.3	13.3	15.2

The gain matching procedure was established using the average scaling factor acquired during mixing speech and noise for PMC compensation. Notably, at 20 dB, the performance of PMC demonstrated comparability to the baseline. However, a discernible degradation in performance was observed as the signal-to-noise ratio (SNR) was decreased. On average, a noticeable enhancement in performance is evident when comparing standard recognition to PMC recognition. This trend persists until reaching 10.2 dB, at which point the performance value of PMC

converges with that of standard recognition. The Detection Error Tradeoff (DET) graphs visually represent a substantial improvement in the recognition trend when PMC is implemented, surpassing the performance of each standard technique in isolation. Figure 1 illustrates the system's performance both with and without the PCM method, providing a graphical representation of the impact of PMC on system performance under varying SNR conditions.

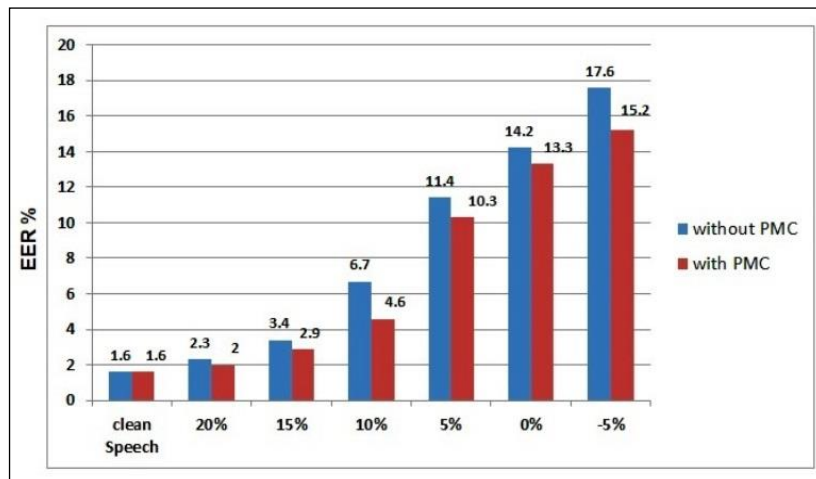


Figure 3 shows the system performance using PMC.

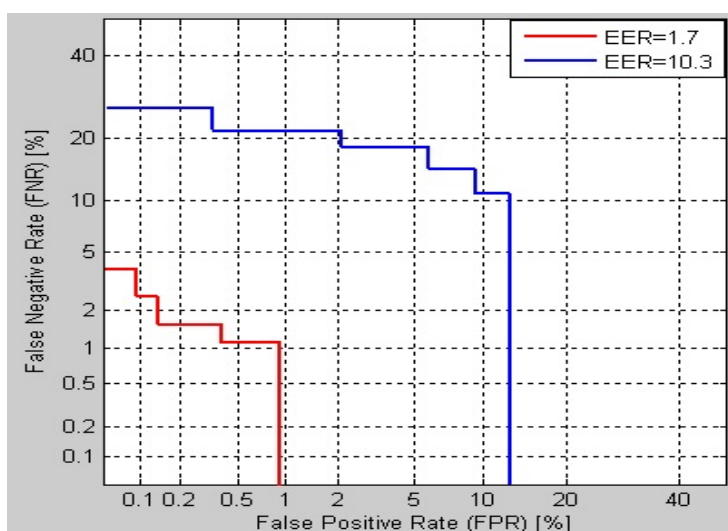


Figure 4 DET plot for clean and SNR=20% using PMC

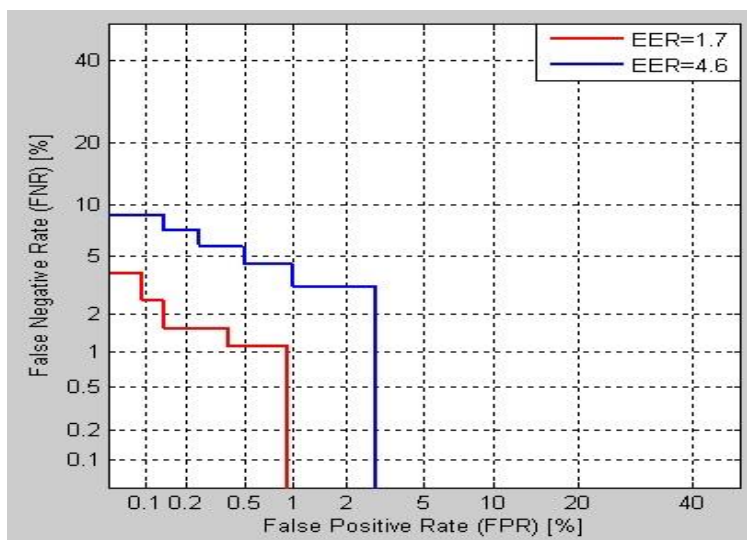


Figure 5 DET plot for clean and SNR=20% using PMC

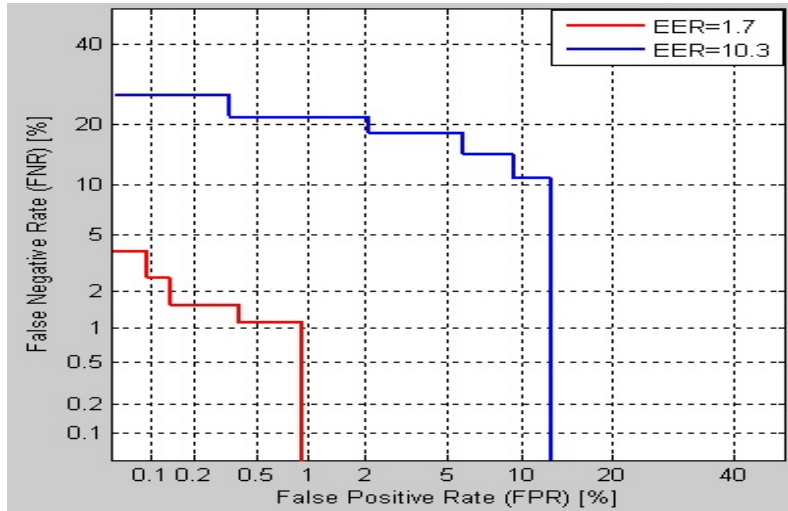


Figure 6 DET plot for clean and SNR=5% using PMC

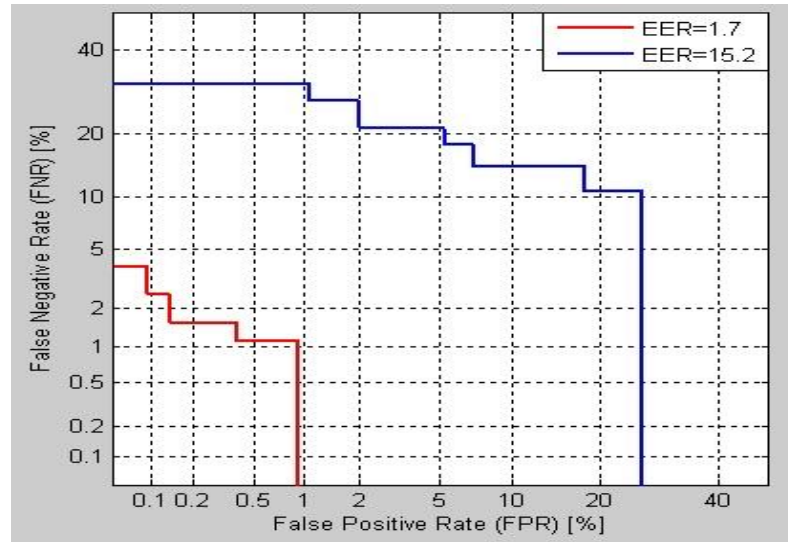


Figure 7 DET plot for clean and SNR=-5% using PMC

5 Concluding Remarks

Speaker recognition systems find application across diverse acoustic environments, prompting concerns about reliability and robustness. Addressing the challenge of achieving robust speaker verification in varied implementations is critical to advancing these systems. This paper investigates the efficacy of parallel model combination (PMC) in enhancing the noise robustness of the Hidden Markov Model (HMM), which is based on text-dependent speaker verification. The experiments utilize speech data sourced from the SALU-AC corpus. The baseline Equal Error Rate (EER) of 1.6% for clean speech experiences a notable degradation when subjected to varying noise levels. For instance, the clean speech EER of 1.7% substantially drops to 11.4% and 14.2% at 5 dB and 0 dB signal-to-noise ratio (SNR); however, the incorporation of PMC results in performance improvement. The performance degradation is notably reduced to 4.6%, 10.3%, 13.3%, and 15.2% at 10 dB, 5 dB, 0 dB, and -5 dB SNR, showcasing the effectiveness of PMC in mitigating the impact of noise. This enhancement in authentication performance becomes particularly evident when the SNR falls below 10 dB. These findings underscore the potential of PMC as a valuable tool for bolstering the robustness of speaker verification systems in challenging acoustic conditions.

Authors' Contributions: All authors have worked equally to write and review the manuscript.

Data Availability Statement: All data are presented in the paper, along with their related references, in the Application section.

Conflicts of Interest: The authors declare no conflict of interest.

Funding: There are no funds received for this paper.

Use of Generative-AI tools declaration: The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

References

- [1] K. A. Al-Karawi, A. H. Al-Noori, F. F. Li, and T. Ritchings, "Automatic Speaker Recognition System in Adverse Conditions--Implication of Noise and Reverberation on System Performance," *International Journal of Information and Electronics Engineering*, vol. 5, pp. 423-427, 2015.
- [2] M. M. Abdelwahab, K. A. Al-Karawi, E. Hasanin, and H. Semary, "Autism spectrum disorder prediction in children using machine learning," *Journal of Disability Research*, vol. 3, pp. 1-9, 2024.
- [3] J. P. Campbell, "Speaker recognition: a tutorial," *Proceedings of the IEEE*, vol. 85, pp. 1437-1462, 1997.
- [4] K. A. Al-Karawi and D. Y. Mohammed, "Improving short utterance speaker verification by combining MFCC and Entropy in Noisy conditions," *Multimedia Tools and Applications*, vol. 80, pp. 22231-22249, 2021.
- [5] K. A. Al-Karawi and D. Y. Mohammed, "Using combined features to improve speaker verification in the face of limited reverberant data," *International Journal of Speech Technology*, vol. 26, pp. 789-799, 2023.
- [6] K. A. Al-Karawi and F. Li, "Robust speaker verification in reverberant conditions using estimated acoustic parameters—A maximum likelihood estimation and training on the fly approach," in *2017 Seventh International Conference on Innovative Computing Technology (INTECH)*, 2017, pp. 52-57.
- [7] K. A. Al-Karawi and S. T. Ahmed, "Model selection toward robustness speaker verification in reverberant conditions," *Multimedia Tools and Applications*, vol. 80, pp. 36549-36566, 2021.
- [8] K. A. Al-Karawi, "Face mask effects on speaker verification performance in the presence of noise," *Multimedia Tools and Applications*, vol. 83, pp. 4811-4824, 2024.
- [9] N. Alotaibi, I. Al-Dayel, I. Elbatal, M. R. Ezzeldin, M. Elgarhy, and K. A. Al-karawi, "STATISTICAL ANALYSIS OF COVID-19 PANDEMIC IN SAUDI ARABIA," *Advances and Applications in Statistics*, vol. 74, pp. 107-118, 2022.
- [10] K. A. Yousif, "Characters and Digits Recognition Using Neural Network Learned by Particle Swarm Optimization," *Diyala Journal For Pure Science*, vol. 10 pp. 73-91, 2014.
- [11] K. A. Al-Karawi, "Mitigate the reverberation effect on the speaker verification performance using different methods," *International Journal of Speech Technology*, vol. 24, pp. 143-153, 2021/03/01 2021.
- [12] A. Aljuboory, L. A. Tawfeeq, and K. A. Al-Karawi, "Pushing towards ehealth for iraqi hypertensives: an integrated class association rules into SECI model," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, pp. 522-533, 2021.
- [13] M. M. Abdelwahab, K. A. Al-Karawi, and H. E. Semary, "Deep learning-based prediction of Alzheimer's disease using microarray gene expression data," *Biomedicines*, vol. 11, p. 3304, 2023.
- [14] D. A. Reynolds, "Speaker identification and verification using Gaussian mixture speaker models," *Speech communication*, vol. 17, pp. 91-108, 1995.
- [15] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital signal processing*, vol. 10, pp. 19-41, 2000.
- [16] A. S. Alenizi and K. A. Al-karawi, "Cloud Computing Adoption-Based Digital Open Government Services: Challenges and Barriers," in *Proceedings of Sixth International Congress on Information and Communication Technology*, 2022, pp. 149-160.
- [17] K. A. Al-Karawi and D. Y. Mohammed, "Early reflection detection using autocorrelation to improve robustness of speaker verification in reverberant conditions," *International Journal of Speech Technology*, vol. 22, pp. 1077-1084, 2019.
- [18] K. A. Al-Karawi and B. Al-Bayati, "The effects of distance and reverberation time on speaker recognition performance," *International Journal of Information Technology*, vol. 16, pp. 3065-3071, 2024.
- [19] J. Ortega-García and J. González-Rodríguez, "Overview of speech enhancement techniques for automatic speaker recognition," in *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, 1996, pp. 929-932.
- [20] A. S. Alenizi and K. A. Al-karawi, "Machine learning approach for diabetes prediction," in *International Congress on Information and Communication Technology*, 2023, pp. 745-756.

- [21] A. P. Varga and R. K. Moore, "Hidden Markov model decomposition of speech and noise," in *International Conference on Acoustics, Speech, and Signal Processing*, 1990, pp. 845-848.
- [22] J. Ming, T. J. Hazen, J. R. Glass, and D. A. Reynolds, "Robust speaker recognition in noisy conditions," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, pp. 1711-1723, 2007.
- [23] M. J. Gales and S. J. Young, "Robust continuous speech recognition using parallel model combination," *IEEE Transactions on Speech and Audio Processing*, vol. 4, pp. 352-359, 1996.
- [24] A. H. Al-Noori, K. A. Al-Karawi, and F. F. Li, "Improving Robustness of Speaker Recognition in Noisy and Reverberant Conditions via Training," in *Intelligence and Security Informatics Conference (EISIC), 2015 European*, 2015, pp. 180-180.
- [25] M. M. Abdelwahab, K. A. Al-Karawi, and H. Semary, "Integrating gene selection and deep learning for enhanced Autisms' disease prediction: a comparative study using microarray data," *AIMS Mathematics*, vol. 9, pp. 17827-17846, 2024.
- [26] K. A. Al-karawi, "Internet of Things (IoT) about Disabilities: Disabilities in relation to the Internet of Things (IoT)," *ScienceOpen Preprints*, 2023.
- [27] A. S. Alenizi and K. A. Al-Karawi, "Internet of things (IoT) adoption: challenges and barriers," in *Proceedings of Seventh International Congress on Information and Communication Technology: ICICT 2022, London, Volume 3*, 2023, pp. 217-229.
- [28] d. Mohammed, K. A. Al-Karawi, P. Duncan, and F. F. Li, "Overlapped Music segmentation using a new Effective Feature and Random Forests," *International Journal Of artificial intelligence (IN-IA)*, vol. 8, pp. 181-189, June 2019.
- [29] A. S. Alenizi and K. A. Al-Karawi, "Effective biometric technology used with big data," in *Proceedings of Seventh International Congress on Information and Communication Technology: ICICT 2022, London, Volume 3*, 2022, pp. 239-250.
- [30] K. A. Al-Karawi, "Real-time adaptive training for forensic speaker verification in reverberation conditions," *International Journal of Speech Technology* vol. 26, pp. 1079-1089, 2023.
- [31] F. Al-Dhaher, D. Y. Mohammed, M. Khalaf, K. Al-Karawi, M. Sarfraz, and M. M. Al Maathidi, "Real-Time Lie-Speech Determination Using Voice-Stress Technology," *Iraqi Journal For Computer Science and Mathematics*, vol. 5, pp. 81-93, 2024.
- [32] L. P. Wong and M. Russell, "Text-dependent speaker verification under noisy conditions using parallel model combination," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, 2001, pp. 457-460.
- [33] H. Semary, K. A. Al-Karawi, M. M. Abdelwahab, and A. Elshabrawy, "A Review on Internet of Things (IoT)-Related Disabilities and Their Implications," *Journal of Disability Research*, vol. 3, pp. 1-16, 2024.
- [34] K. A. Al-Karawi and A. S. Alenizi, "Reverberation Time and Distance Impact on the Equal Error Rate," in *Proceedings of Ninth International Congress on Information and Communication Technology: ICICT 2024, London, Volume 10*, 2024, p. 13.
- [35] A. S. Alenizi and K. A. Al-Karawi, "Speaker Recognition with Deep Learning Approaches: A Review," in *International Congress on Information and Communication Technology*, 2024, pp. 481-499.
- [36] H. Semary, K. A. Al-Karawi, and M. M. Abdelwahab, "Using voice technologies to support disabled people," *Journal of Disability Research*, vol. 3, pp. 1-8, 2024.
- [37] K. A. Al-karawi, "Real-time adaptive training for forensic speaker verification in reverberation conditions," *International Journal of Speech Technology*, pp. 1-11, 2023.
- [38] M. Hamidi, H. Satori, O. Zealouk, K. Satori, and N. Laaidi, "Interactive voice response server voice network administration using hidden markov model speech recognition system," in *2018 Second World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, 2018, pp. 16-21.
- [39] R. Singh, B. Raj, and R. M. Stern, "Model compensation and matched condition methods for robust speech recognition," in *Noise reduction in speech applications*, ed: CRC press, 2018, pp. 245-275.
- [40] L. Docio-Fernandez and C. Garcia-Mateo, "Noise model selection for robust speech recognition," in *Fifth International Conference on Spoken Language Processing*, 1998.
- [41] K. Al-Karawi, "Robust speaker recognition in reverberant condition-toward greater biometric security," University of Salford, 2018.