University of Baghdad
College of Engineering
JE
JOURNAL OF ENGINEERING

Journal of Engineering
journal homepage: www.jcoeng.edu.iq
Volume 32 Number 3 March 2026

A Dual-Stage Perceptual-Harmonic Hybrid Estimator for Speech Enhancement

Sally Taha Yousif , Basheera M. Mahmmod *

Department of Computer Engineering, College of Engineering, University of Baghdad, Baghdad, Iraq

ABSTRACT

This paper proposes a hybrid speech enhancement estimator that integrates the Perceptually-motivated Karhunen–Loève Transform (PKLT) with the Dual-Masking Harmonic-based (DMH) algorithm in a unified framework termed PKDMH. The main novelty lies in combining perceptual subspace projection with harmonic-residual suppression, enabling the system to jointly remove noise while preserving speech-relevant spectral cues. PKLT first performs perceptual subspace projection and suppresses inaudible components, after which DMH eliminates remaining broadband and harmonic residuals. The proposed PKDMH system was evaluated using the TIMIT dataset contaminated with five noise types: White, Pink, F16, Airport, and Car noise—across five SNR levels (−10 dB, −5 dB, 0 dB, +5 dB, +10 dB). Objective evaluation used the standard perceptual and signal-level measures of PESQ, STOI, SNRseg, Csig, Cbak and Covl. Results show that the enhanced quality of separation and speech signal ratio between enhanced signals and original target binary mask cause obvious improvements in quantity, with average PESQ gains of 1.099, 0.888 and 0.824 for White, Pink and F16 noise, respectively. These results bring out the subjective benefit of the PKDMH cascade, in terms of being a more robust enhancement approach under low SNR and acoustically varying cases.

Keywords: DMH, Intelligibility, PKLT, PKDMH, Quality metrics, Speech enhancement.

1. INTRODUCTION

Speech Enhancement (SE) is a fundamental component of modern Digital Signal Processing (DSP). Its primary objective is the reconstruction of clean speech from noisy observations. SE plays a critical role in applications such as automatic speech recognition (ASR), telecommunication systems, voice-controlled interfaces, hearing-assistive devices, and human–computer interaction **(Loizou, 2013; Vary and Martin, 2006)**. The quality of speech-based systems is commonly rated in two main categories: Speech Quality and Speech Intelligibility. Speech quality is indicative of how clearly and naturally it is spoken, whereas intelligibility refers to how accurately the listener can translate speech into words **(Kolbæk et al., 2016; Jerjees et al., 2023)**.

*Corresponding author

Peer review under the responsibility of University of Baghdad.
https://doi.org/10.31026/j.eng.2026.03.10

Article received: 02/09/2025
Article revised: 20/01/2026
Article accepted: 30/01/2026
Article published: 01/03/2026

Quality of speech refers to the clarity and understandability of a speech signal to its intended listeners. It demonstrates how nice the sound is, to what extent it's clear of distortion, and it lacks artefacts in good acoustic environments. Speech intelligibility, meanwhile, refers to the listener's capability of accurately recognizing and comprehending spoken language. Both of these features are crucial for evaluating the effectiveness of spoken communication **(Elert, 2016; Jerjees et al., 2023).** In real environments, speech signals are frequently degraded by different types of additive noise. These include white, pink, F16, factory, and Buccaneer noise. Each type masks phonetic cues and reduces both intelligibility and perceptual quality **(Hu and Loizou, 2007; Soon et al., 1998).**

To address this challenge, SE algorithms are widely used across several domains. Hearing aids and cochlear implants rely on SE to suppress background noise and enhance speech cues. ASR systems and smart assistants employ both classical signal-processing and DNN-based approaches to improve robustness under adverse acoustic conditions **(Ochieng, 2023).** Telecommunication systems use SE to maintain clarity in low-bandwidth or noisy channels. In addition, forensic analysis frequently utilizes enhancement techniques to recover speech from degraded recordings. Despite the increasing adoption of deep neural networks, recent findings indicate that data requirements, computational costs, and generalization limitations still constrain their use in real-time and low-resource scenarios **(Al-Khateeb et al., 2025; Yousif et al., 2025)**.

Many people have used traditional speech enhancement methods like spectral subtraction **(Verteletskaya and Simak, 2010; Boll, 1979; Stylianou, 2001)**, Wiener filtering **(Scalart and Filho, 1996; Xia and Bao, 2014),** Kalman filtering **(Nabi et al., 2016; Roy et al., 2021),** and Minimum Mean Square Error (MMSE) estimators **(Ephraim and Malah, 1984; Hasan and Hasan, 2010; Shi et al., 2023)** have been extensively studied and applied. These methods are effective under stationary noise conditions. However, their performance often degrades when noise is highly non-stationary. Moreover, they may introduce artifacts such as musical noise or over-smoothing, which degrade perceptual quality **(Loizou, 2013)**. Several refinements have been proposed to mitigate these limitations. These include perceptual models **(Jabloun and Champagne, 2003)**, improved noise estimation **(Cohen, 2002),** and decision-directed a priori SNR estimation. Nevertheless, the trade-off between noise reduction and speech distortion remains a persistent challenge.

Recent work on speech enhancement has shifted towards deep-learning-based and hybrid signal-processing architectures aimed at addressing the highly non-stationary nature of noise and to enhance perceptual quality beyond traditional statistical-based methods. Among them, some of the prominent new directions are phase-aware complex-domain models, perceptually motivated deep filtering and metric-based learning strategies that directly optimize the perceptual and intelligibility-related objectives **(Hu et al., 2020; Fu et al., 2021).** Despite their strong performance, these approaches often suffer from high data dependency, limited generalization to unseen noise conditions, and increased computational complexity, which motivates the development of robust hybrid designs that preserve interpretability while benefiting from modern learning-based techniques **(Michelsanti et al., 2021).**

Recent research has increasingly focused on hybrid enhancement frameworks that combine the strengths of different algorithmic paradigms **(Mahmmod et al., 2021).** The Perceptually-motivated Karhunen–Loève Transform (PKLT) has shown promising performance in attenuating perceptually irrelevant noise by exploiting auditory masking and subspace modeling **(Jabloun and Champagne, 2003).** Meanwhile, harmonic-based estimators such as the Dual-Masking Harmonic (DMH) method are effective in suppressing residual and harmonic noise patterns while preserving voiced-speech structure **(Plapous et al., 2006).**

This work proposes a novel two-stage hybrid speech enhancement framework, termed PKDMH, that employs the cascaded combination of perceptually motivated subspace processing and harmonic-aware masking. In contrast to previous hybrid approaches, its implementation combines PKLT for psycho-acoustically subspace-projecting perceptually irrelevant noise components with DMH for post-elimination of remaining residual and harmonic noise artefacts that commonly result from applying a filtering strategy based on subspace projection. The cascaded design provides complementary noise suppression while preserving speech structure, facilitating simultaneous improvements in speech quality and intelligibility without resorting to large training data or high complexity. Experimental results on different noise types and SNR levels have verified that the proposed PKDMH framework is superior to classical and state-of-the-art enhancement methods.

2. PRELIMINARY

This section describes in detail the two main SEA (DMH and PKLT) as well as their mathematical equations used for computing the enhanced speech signal that will be used in the proposed work, as follows:

2.1 Perceptually-Motivated Karhunen-Loève Transform (PKLT)

The PKLT algorithm, proposed by **(Jabloun and Champagne; 2003),** applies signal subspace analysis enhanced with psychoacoustic principles to suppress perceptually irrelevant noise. The noisy speech is decomposed into orthogonal eigen components, and only components exceeding the auditory masking threshold are retained. Voice Activity Detection (VAD) is applied before PKLT processing to identify non-speech frames for noise covariance estimation. A conventional energy-based VAD is employed, where frames with energy below a predefined threshold are classified as noise-only frames. Each noisy speech frame is windowed and represented as a column vector $y(n)$, from which the short-time covariance matrix is computed as:

$$R_y = E[y(n)y^H(n)] \tag{1}$$

where $E[\cdot] denotes\ statistical\ expectation$, $y^H(n)$is the Hermitian conjugate of $y(n)$. Eigen-decomposition is then applied to the covariance-matrix **(Ephraim and Van Trees, 1995):**

$$R_y = U\Lambda U^H \tag{2}$$

$R_y \in R^{L\times L}$is the noisy speech covariance matrix, $U \in R^{L\times L}$is the eigenvector matrix, and $\Lambda = diag(\lambda_1, \dots, \lambda_L)$is the diagonal matrix of eigenvalues sorted in descending order. The eigenvectors associated with larger eigenvalues span the signal subspace, while the remaining eigenvectors correspond to the noise subspace. This decomposition is essentially the Karhunen–Loève Transform (KLT), which represents the signal in terms of uncorrelated orthogonal basis functions obtained from the covariance matrix. In practice, KLT concentrates most of the structured speech energy into a few dominant dimensions, while distributing the unstructured noise energy more evenly across all dimensions. This makes it possible to selectively retain the signal subspace and attenuate the noise subspace, providing a compact and adaptive representation that is particularly effective for speech enhancement **(Huang and Zhao, 2000; Vetter, 2001).**

A psychoacoustic model, such as that proposed by **(Zwicker and Fastl,1999; Zwicker and Fastl, 1990),** estimates the perceptual masking threshold for each frequency:

$$M(f) = T_q(f) + \Delta \tag{3}$$

where $T_q(f)$ is the absolute threshold of hearing and Δ is a correction factor for critical band integration. Only eigencomponents with energy above this threshold are retained.

In PKLT, the perceptual masking threshold $M(f)$ is included implicitly by means of a spectral decomposition of the covariance matrix, thus allowing for a mapping between eigenvectors and frequency bins to be done through spectral shaping. Eigenvalues correspond to the energy of broadband components at lower energies the latter components falling beneath a masking threshold are suppressed when representing perceptually important speech and attenuating inaudible noise.

The enhanced signal $\hat{s}(n)$ is reconstructed by projecting the noisy signal onto the perceptual subspace:

$$\hat{s}(n) = U_p U_p^H y(n) \tag{4}$$

where U_p is the submatrix of U containing only eigenvectors corresponding to components above the perceptual masking threshold. This process ensures that masked or inaudible noise is suppressed while preserving the intelligibility and quality of the speech signal. PKLT, using a binary subspace projection for clarity, the practical implementation employs a soft-decision gain function. Each eigen component is weighted according to its signal-to-noise dominance rather than being strictly kept or discarded. Specifically, for the k-th eigencomponent, the gain is defined as:

$$G_k = max\ \left(0,\ 1 - \frac{\lambda_{v,k}}{\lambda_k}\right), \tag{5}$$

where $\lambda_k$and $\lambda_{v,k}$denote the noisy-signal and noise eigenvalues, respectively. The enhanced signal is then reconstructed using a diagonal gain matrix, which provides smooth attenuation of noise-dominated components and reduces artefacts compared to hard masking **(Ephraim and Van Trees, 1995)**.

2.2 Dual-Masking Harmonic-based Method (DMH)

The Dual-Masking Harmonic-based (DMH) method is a significant method in SE. This technique is based on the framework developed by **(Plapous et al., 2006),** which combines Two-Step Noise Reduction (TSNR) with a Harmonic Regeneration Noise Reduction (HRNR) strategy. The method is specifically designed to address both stationary and non-stationary residual noise components, particularly harmonic distortions that remain after the first-stage enhancement. This nonlinear system efficiently regenerates the degraded harmonics of distorted speech. The resulting synthetic signal is produced to improve the a priori signal-to-noise ratio (SNR). It is based on two steps as presented in the following sections.

2.2.1 Gain estimation based on TSNR

TSNR method strengthens the traditional decision-directed algorithm, which has a built-in frame delay bias. To measure it, a posteriori Signal-to-Noise Ratio (SNR) is computed as follows:

$$\gamma(k) = |Y(k)|^2 / \lambda_d(k) \tag{6}$$

$Y(k)$ is the noisy speech spectrum and $\lambda_d(k)$ indicates estimated noise power spectral density. The a priori SNR is then calculated via a modified decision-directed estimator:

$$\xi(k) = \alpha \cdot |\hat{S}(k-1)|^2 / \lambda_d(k) + (1-\alpha) \cdot max[\gamma(k) - 1, 0] \tag{7}$$

where α is athe smoothing factor in the TSNR decision-directed estimator is set to $\alpha = 0.97$, providing stable a priori SNR estimation. This refined estimate is used to derive the TSNR-based spectral gain:

$$G_{TSNR}(k) = \xi(k) / (1 + \xi(k)) \tag{8}$$

where $k$denotes the frequency-bin index, $Y(k)$is the short-time Fourier transform (STFT) of the noisy speech signal, $\lambda_d(k)$is the estimated noise power spectral density, $\hat{S}(k-1)$is the enhanced speech spectrum from the previous frame, and $\alpha \in [0,1]$is a smoothing factor controlling the temporal averaging in the decision-directed estimator.
This gain is then applied to the noisy spectrum, curbing additive noise while reducing musical artifacts and preserving meaning.

2.2.2 Harmonic-Residual Noise Reduction (HRNR)

HRNR eliminates the residual noise component associated with voiced speech. A straightforward nonlinear operation in the time domain, such as half-wave rectification, is applied to the enhanced speech output from TSNR's first stage to produce an artificial, harmonic-rich signal:

$$y_{\mathrm{harm}}(n) = \max\{ y_{\mathrm{TSNR}}(n), 0 \} \tag{9}$$

Where:

- $y_{\mathrm{TSNR}}(n)$: time-domain output of the TSNR stage,
- $y_{\mathrm{harm}}(n)$: rectified signal containing enhanced harmonic energy.

This easy nonlinearity actually creates higher harmonics (integer-multiples of the main frequency), which were partly lost during the first filtering.
This cleaned-up signal then becomes part of the a priori SNR, mixing TSNR's estimate power with the reconstructed harmonic.

$$\xi_{HRNR}(k) = \rho(k) \cdot |\hat{S}_{TSNR}(k)|^2 + (1 - \rho(k)) \cdot |S_{harmo}(k)|^2 \tag{10}$$

where $\hat{S}_{TSNR}(k)$is the TSNR-enhanced speech spectrum, $S_{harmo}(k)$is the harmonic spectrum obtained from nonlinear time-domain processing followed by spectral analysis, and $\rho(k)$is a frequency-dependent mixing coefficient defined as the TSNR gain $G_{TSNR}(k)$.The final HRNR gain is calculated as follows:

$$G_{HRNR}(k) = \xi_{HRNR}(k) / (1 + \xi_{HRNR}(k)) \tag{11}$$

This gain is then multiplied by the original noisy spectrum to obtain a cleaner signal:

$$\hat{S}(k) = G_{HRNR}(k) \cdot Y(k) \tag{12}$$

Finally, the enhanced time domain signal is found using inverse, giving speech with residual noise removed and preserving its harmonic structures.

2.3 The Proposed Algorithm

Single-stage speech enhancement methods suffer from inherent limitations when operating under diverse and highly degraded acoustic conditions. The PKLT algorithm is effective in suppressing perceptually irrelevant noise through subspace projection and auditory masking; however, it often leaves residual and harmonic noise components, particularly in voiced segments and at low SNR levels. Conversely, the DMH algorithm efficiently suppresses residual and harmonic noise but relies on the availability of a reasonably noise-reduced input signal to operate reliably. This complementary behavior motivates the proposed cascaded PKLT–DMH architecture, which aims to jointly improve speech quality, intelligibility, and residual noise suppression. In the proposed PKDMH framework, PKLT is first applied to remove inaudible and perceptually insignificant noise components while preserving the dominant speech structures. The output of PKLT is then processed by DMH to further suppress residual and harmonic noise artifacts that remain after subspace-based filtering.

To investigate the impact of processing order, both enhancement algorithms were initially implemented independently. Subsequently, two cascade configurations—PKLT→DMH and DMH→PKLT—were evaluated under identical experimental conditions. This self-comparison allows a systematic assessment of the effect of algorithm ordering on enhancement performance. Experimental results consistently show that the use of PKLT before DMH gives less background distortion and better speech quality and intelligibility, in particular at low-SNR and time-varying noise conditions. Accordingly, the PKLT→ DMH cascade is chosen as our final proposed configuration, and we evaluate its robustness and effectiveness versus state-of-the-art approaches.

2.4 The Proposed Work

The proposed work is composed of two stages, each with its own function. The PKLT module is the first stage in the proposed hybrid speech architecture for getting rid of noise. The orthogonal eigenvectors (basis functions) of the noisy speech signal are derived from the eigen-decomposition of the signal's short-time correlation matrix. This signal subspace architecture is where PKLT works. The key point is that in this representation, speech and noise exist in different subspaces. Noise energy is more evenly spread out throughout the other dimensions, while speech signal energy is usually focused on a smaller group of high-energy eigen components. PKLT gets rid of noise components that are below the auditory masking threshold by finding and keeping just the eigenvectors that are associated to perceptually important components. According to **(Jabloun and Champagne, 2003).** Psychoacoustic modeling is used to discover the point at which sounds are blocked by other sounds. This idea is based on the fact that the human ear can't hear some weak noises when there are stronger signals at nearby frequencies. PKLT guarantees that only eigen components that are both energetic and audible to the human ear are kept by combining both perceptual needs into the subspace filtering process. The algorithm can get rid of noise that isn't needed for speech without affecting vital parts of speech. At this point, a Voice Activity Detection (VAD) **(Scalart and Filho, 1996)** module is included to tell the difference between parts that are speech-active and parts that are not. To fully confirm the separation of the signal subspace and noise subspace, the noise covariance matrix must be accurately calculated using non-speech frames. The noisy speech is divided into overlapping frames, placed in windows, and transformed into column vectors to enable the computation of covariance subsequent to the determination of noise statistics. After eigen-decomposition, the components that are preserved go via a gain function that is based on psychoacoustic masking thresholds. After that, the better signal from the subspace is put back together. The initial step in this process maintains speech parts with as little distortion as possible and cuts down on noise that can't be heard. There may still be problems that do not go away, such low-level "musical noise" and harmonic distortions, especially in voiced areas and when the SNR is low. The Dual-Masking Harmonic-based (DMH) module obtains its output from the PKLT because it reduces targeted residual noise and restores harmonics.

We use Harmonic Regeneration Noise Reduction (HRNR) and Two-Step Noise Reduction (TSNR) in the DMH stage. TSNR enhances the decision-directed (DD) **(Ephraim and Malah 1984; Cappe, 1994)** a priori SNR estimation technique, which is effective. However, suffers from issues such as frame-delay bias, sluggish adaptation to rapidly fluctuating noise, and the risk of underestimating SNR during speech onsets. These could result in either sounding weird while speech or not cutting down on noise enough for short-term frames. To address these limitations, we add a phase term that incorporates the ramp-up time and allows us to learn more effectively across different noise patterns. This improves the SNR estimation. Afterwards, HRNR restores any harmonic component that could be lost or made weaker during the initial processing. An artificial harmonic-enhanced signal is produced by HRNR using nonlinear time-domain techniques including half-wave rectification.

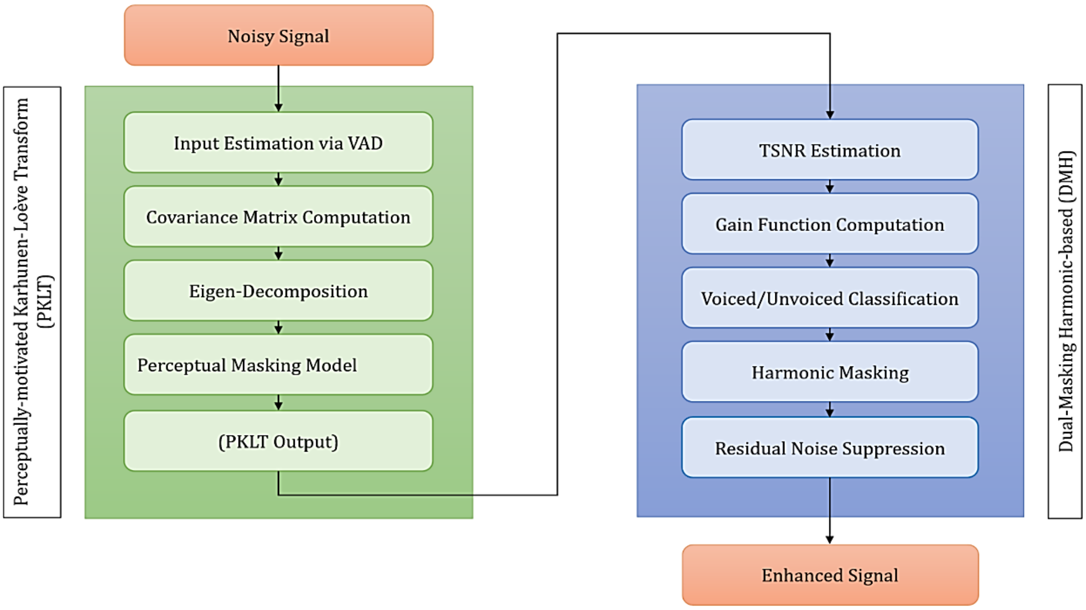

Noisy Signal
Perceptually-motivated Karhunen-Loève Transform (PKLT)
Input Estimation via VAD
Covariance Matrix Computation
Eigen-Decomposition
Perceptual Masking Model
(PKLT Output)
TSNR Estimation
Gain Function Computation
Voiced/Unvoiced Classification
Harmonic Masking
Residual Noise Suppression
Dual-Masking Harmonic-based (DMH)
Enhanced Signal

Figure 1. Proposed Two-Stage Hybrid Speech Enhancement System.

This recovers the harmonic content that is important for quality and intelligibility in speech. Subjectively, the voice sounds more natural and sonorous through this addition of TSNR output with the regenerated harmonic components. PKDMH method uses perceptually directed subspace filtering from PKLT and targeted residual noise reduction and harmonic reconstruction from DMH. This strikes a nice compromise between noise cancellation and speech clarity, even in crowded areas.

3. RESULTS AND DISCUSSION

This section will cover two types of comparison. The first one involves a self-comparison process to analyze the performance and determine the overall value of the proposed work. Then, a comparison with other related work is made that evaluates the current system against previous approaches to highlight its strengths and weaknesses. Experiments were conducted using the TIMIT speech corpus **(Garofolo et al., 1993)**, which contains clean, phonetically balanced speech recordings. Clean utterances were artificially corrupted by adding five types of noise—White, Pink, Car, Airport, and F16 at SNR levels of 10 dB, 5 dB, 0 dB, -5 dB, and −10dB. Objective quality is assessed using PESQ, SNRseg, Cbak, and Covl **(Hu and Loizou, 2007; Rix et al., 2001),** while intelligibility is measured through STOI and PESQ **(Taal et al., 2011).**

Four configurations are examined in a self-comparison study: PKLT-SEA only, DMH-SEA only, DMH→PKLT, and PKLT→DMH. All configurations are tested under identical experimental setups using the same dataset, noise types, and SNR levels. Results indicate that PKLT→DMH consistently outperforms the other configurations across all objective metrics, PESQ, STOI, Cbak, Covl, and Csig at SNR levels of (10 dB, 5 dB, 0 dB, −5 dB, and -10 dB), achieving performance comparable to or better than recent deep learning-based approaches such as CNN and FCNN **(Al-Zubaidi et al., 2024)**, particularly in challenging noise environments.

In both stages of the proposed system, the noisy speech signal is first sampled at 16 kHz and segmented into overlapping frames of 20 ms duration, corresponding to 320 samples per frame, with a 50% overlap between consecutive frames. Each frame is multiplied by a Hann window before transformation, as it effectively reduces spectral leakage and improves spectral estimation accuracy **(Harris, 1978)**. This windowing strategy ensures smooth transitions between adjacent frames and maintains spectral consistency, which is essential for both the PKLT and DMH processing stages.

3.1 The Proposed System's Self-Evaluation

Different scenarios have been implemented to analyze the proposed SEA by using different quality and intelligibility measures. A self-comparison between normal individual cases and dual hybrid cases is performed. Three objective quality metrics are used to evaluate the suggested hybrid speech enhancement system: PESQ (Perceptual Evaluation of Speech Quality), SNRseg (Segmental Signal-to-Noise Ratio), and Cbak (Background Intrusiveness) across three different noise types: F16, Pink, and White. Besides, five SNR levels−10 dB, −5 dB, 0 dB, 5 dB, and 10 dB—are used to test each noise condition. In this experiment, the noisy (unprocessed) signal has been compared to four types of enhanced signals based on four types of estimators to show the rate of improvement for each one. These four estimator algorithms are PKLT **(Jabloun and Champagne ;2003),** DMH **(Plapous et al., 2006),** the cascade estimator of DMH→PKLT, and finally the cascade estimators of PKLT→DMH where the results of each processing estimators have been calculated. The comprehensive findings are arranged in three tables: **Table 1** displays PESQ scores, **Table 2** displays Cbak scores, and **Table 3** displays SNRseg values.

Table1. PESQ for each algorithm for three types of noise with five different levels of SNR**.**

Results of white noise					
SNR	**Noisy**	**PKLT**	**DMH**	**DMH+PKLT**	**PKLT+DMH**
-10	0.7361	1.1548	1.2989	1.5900	**1.6000**
-5	0.9793	1.4807	1.6902	2.0091	**2.0091**
0	1.3112	1.7684	2.1168	2.3626	**2.3627**
5	1.6867	2.1452	2.5331	2.7357	**2.7358**
10	2.0752	2.5221	2.8917	3.0407	**3.0407**
Results of F16 noise					
SNR	**Noisy**	**PKLT**	**DMH**	**DMH+PKLT**	**PKLT+DMH**
-10	0.9018	0.5621	1.3570	1.3643	**1.3644**
-5	1.0954	1.0994	1.7676	1.8517	**1.8518**
0	1.4507	1.5426	2.1795	2.2900	**2.3000**
5	1.8150	2.0234	2.5246	2.6293	**2.6294**
10	2.1980	2.5009	2.8862	2.9910	**2.9911**
Results of pink noise					
SNR	**Noisy**	**PKLT**	**DMH**	**DMH+PKLT**	**PKLT+DMH**
-10	0.6914	0.8921	1.3057	1.4557	**1.4557**
-5	0.9858	1.3093	1.7392	1.9075	**1.9075**
0	1.3606	1.6910	2.1508	2.2979	**2.2979**
5	1.7637	2.0319	2.5386	2.6666	**2.6666**
10	2.1630	2.5611	2.8906	2.9986	**2.9986**

Table 2. Cbak for each algorithm for three types of noise with five different levels SNR.

Results of white noise					
SNR	**Noisy**	**PKLT**	**DMH**	**DMH+PKLT**	**PKLT+DMH**
-10	1.0370	1.5085	1.5146	1.6606	**1.6607**
-5	1.1853	1.8607	1.8952	2.0519	**2.0520**
0	1.5726	2.2309	2.2754	2.4139	**2.4141**
5	2.0528	2.6429	2.6292	2.7635	**2.7636**
10	2.5587	3.0373	2.9307	3.0547	**3.0548**
Results of F16 noise					
SNR	**Noisy**	**PKLT**	**DMH**	**DMH+PKLT**	**PKLT+DMH**
-10	1.0412	1.0972	1.3298	1.3836	**1.3837**
-5	1.0966	1.3982	1.7314	1.8369	**1.8370**
0	1.4389	1.8295	2.1303	2.2439	**2.2440**
5	1.9395	2.3484	2.4874	2.6024	**2.6025**
10	2.4710	2.8812	2.8295	2.9530	**2.9531**
Results of pink noise					
SNR	**Noisy**	**PKLT**	**DMH**	**DMH+PKLT**	**PKLT+DMH**
-10	1.0185	1.2255	1.3866	1.4381	**1.4381**
-5	1.1227	1.5514	1.7762	1.8457	**1.8457**
0	1.4389	1.9438	2.1443	2.2233	**2.2233**
5	1.9719	2.4375	2.5322	2.6362	**2.6362**
10	2.5017	2.9202	2.8509	2.9517	**2.9517**

The results show that the highest results have been obtained for the (PKLT→DMH) configuration, confirming the superiority of the new hybrid estimator for speech enhancement, where the results show a consistent trend across all noise types and evaluation metrics. The suggested sequence performs noticeably better than both individual algorithms and the alternative cascade (DMH→PKLT) in terms of PESQ scores **Table 1**, which are a good indicator of perceptual quality, particularly at low SNR levels (−10 dB and −5 dB). This reaffirms the advantage of using perceptual subspace reduction to remove undetectable noise before using harmonic-based-enhancement. Where the last one (DMH) can remove the residual noise and musical noise effectively.

Table 2 shows that the proposed (PKDMH) estimator dominates over and over again at each step compared to both PKLT stage and DMH stage in terms of Cbak, which is a measure to evaluate perceptual intrusiveness of background noise. The PKDMH system attenuates residual and background noise components more effectively while maintaining the clarity of the speech, which is reflected by smaller Cbak values for all types of noise. This enhancement can be attributed to the complementary properties of DMH, with its concentration on residual and harmonic structures, and PKLT as a perceptually guided noise eliminator. By employing these mechanisms in tandem, the proposed system significantly attenuates perceptual noise artefacts more than that achievable by any mechanism alone.

Table 3. SNRseg for each algorithm for three types of noise with five different levels of SNR.

Results of white noise					
SNR	**Noisy**	**PKLT**	**DMH**	**DMH+PKLT**	**PKLT+DMH**
-10	-8.9562	**0.3162**	-2.1531	0.1617	0.1618
-5	-7.3619	**1.5169**	-0.6313	1.2190	1.2191
0	-4.7537	**3.2980**	0.6492	2.2507	2.2508
5	-1.3412	**5.2758**	1.7484	3.1602	3.1602
10	2.5354	**7.2651**	2.6705	3.9499	3.9599
Results of F16 noise					
SNR	**Noisy**	**PKLT**	**DMH**	**DMH+PKLT**	**PKLT+DMH**
-10	-8.8856	-1.8462	-2.3950	**-1.0744**	**-1.0744**
-5	-7.2289	-0.6364	-0.7857	0.3676	**0.3677**
0	-4.6158	1.0069	0.4309	1.4897	**1.4898**
5	-1.1817	**3.4880**	1.6431	2.7115	2.7116
10	2.6859	**6.3869**	2.6387	3.7896	3.7896
Results of pink noise					
SNR	**Noisy**	**PKLT**	**DMH**	**DMH+PKLT**	**PKLT+DMH**
-10	-8.8914	-1.5182	-2.2345	**-0.7519**	**-0.7519**
-5	-7.2466	-0.3842	-0.8374	**0.3443**	**0.3443**
0	-4.6076	1.2534	0.3926	**1.3966**	**1.3966**
5	-1.1775	**3.7009**	1.6047	2.8118	2.8118
10	2.6958	**6.3186**	2.5630	3.6240	3.6240

Remarkably, under moderate and high SNR conditions (5dB - 10 dB), the individual PKLT algorithm performed better than the hybrid configurations in multiple instances in the SNRseg results **Table 3.** This is because the PKLT algorithm preserves more of the speech signal's temporal energy structure under cleaner conditions while suppressing noise less aggressively than the full hybrid system. But this also supports the suggested system's design goal, which is to provide reliable performance in extremely noisy settings. The hybrid system, which targets harmonic patterns and uses PKLT to aid in perceptual filtering, is especially well-suited for high-noise environments. This combination strikes a good balance between quality and intelligibility, which is especially useful in low-SNR environments.

The suggested hybrid system is specifically designed to function well in noisy environments. The synergy between PKLT and DMH is most helpful in these situations, where speech quality and intelligibility are significantly reduced. The DMH stage refines the outcome by boosting harmonically structured speech elements, while the PKLT component guarantees strong suppression of masked and low-energy noise components. While SNRseg may be higher for PKLT alone at moderate and high SNRs due to greater energy preservation, this does not reflect performance under adverse conditions.

At low SNR levels, the hybrid cascade consistently provides more stable enhancement by effectively suppressing residual and harmonic noise that PKLT alone cannot fully remove. This leads to improved perceptual quality and intelligibility, even if the corresponding SNRseg gains are moderate. Therefore, the observed SNRseg behavior highlights the system's robustness in challenging noise environments rather than a limitation of the proposed approach. When comparing the single-stage methods with the two-stage suggested hybrid system and found that (PKLT→DMH) exhibits improved robustness and effectiveness in the overall conditions, especially in unfavorable acoustic environments, which are considered the most complex cases.

3.2 Comparative Analysis based on Speech Intelligibility

According to a detailed data analysis report, which mainly considers speech intelligibility obtained from the Short-Time Objective Intelligibility (STOI) measure, we have found that the proposed PKDMH estimator outperforms all other techniques under different noise situations. The hybrid approach consistently produced the highest STOI scores, with 0.84 and 0.91 for airport and car noise types at SNR of 5 dB and 10 dB, respectively, as illustrated in Fig. 2. Both of them are far below these values for classical and composite lattice systems. In particular, among the more complex hybrid models such as DD+TSNR+HRNR, MMSE+DD+TSNR+HRNR, and Log-MMSE +DD+TSNR+HRNR as described in **(Nasir and Abdulmohsin, 2025)**, showed further degradation under difficult conditions, with scores falling as low as 0.44–0.51. **(Hattaraki and Kambalimath, 2024)**, The spectral subtraction approach failed to surpass an STOI score of 0.78. This scheme demonstrates how well the suggested framework maintains crucial speech cues that are necessary for intelligibility, especially in authentic acoustic situations.

Additionally, the suggested PKDMH estimate also exhibits its dominance in the Perceptual Evaluation of Speech Quality (PESQ) results, which represent the overall perceived naturalness and clarity of the speech signal. The PKDMH approach outperformed all competing methods, consistently achieving PESQ scores between 2.25 and 2.85 in both noise scenarios. Interestingly, under moderate noise conditions, the Spectral Subtraction method **(Hattaraki and Kambalimath, 2024)** occasionally yielded higher PESQ values, but it performed worse in STOI, suggesting a trade-off between perceived quality and intelligibility. The proposed hybrid system, however, achieved high scores for both measures throughout, which implies a well-balanced enhancement strategy. These results support the successful combination of PKLT and DMH modules that improve speech robustness as well as audibility and perceptual acceptability, offering a robust and practical solution for noisy environments in which quality accuracy is important.

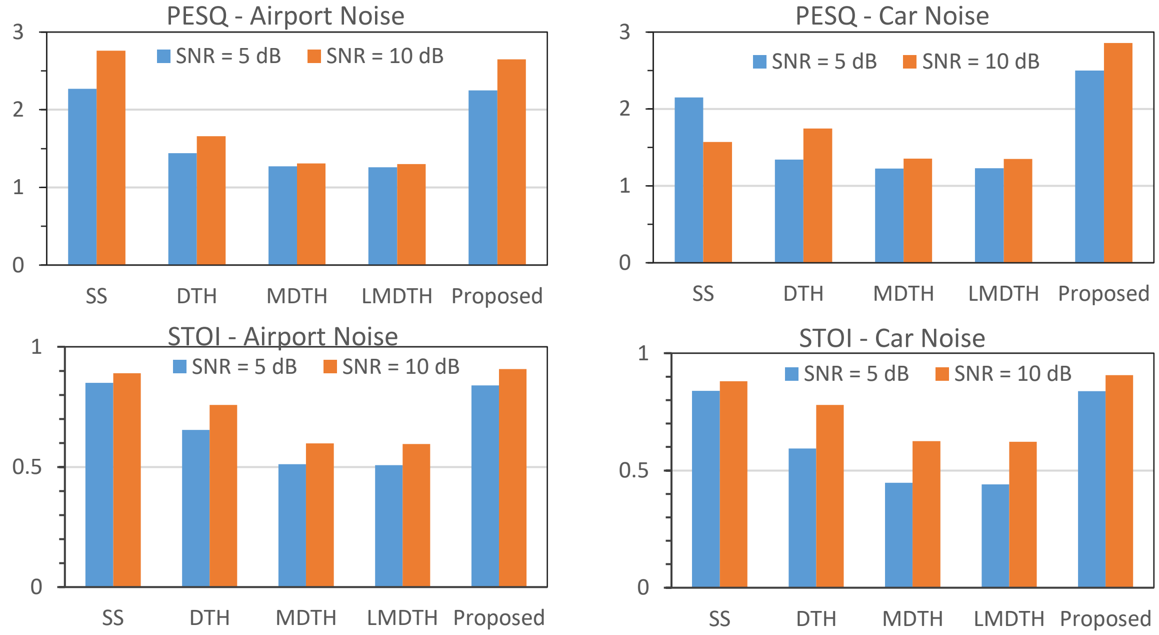


Figure 2. The comparison results of two noise conditions between the proposed algorithm and other SEA for PESQ and STOI. Spectral Subtraction (SS), DD+TSNR+HRNR(DTH), MMSE+DD+TSNR+HRNR(MDTH), Log-MMSE+DD+TSNR+HRNR(LMDTH), Our Hybrid Estimator

3.3 Comparison of Overall Performance Based on Multi-Metric Evaluation

To test the quality of the enhanced signal, four common objective evaluation measures were used in a multi-metric comparison that fully evaluate the efficacy of the suggested hybrid estimator: Perceptual Evaluation of Speech Quality (PESQ), Segmental Signal-to-Noise Ratio (SNRseg), Background Intrusiveness (Cbak), and Overall, Speech Quality (Covl). Three types of noise were used for the comparison purpose: white, pink, and F16 at two critical levels of SNR ratio, 0 dB and -5 dB. Current baseline methods like Gaussian-Gaussian Bayesian models and Laplace-Laplace **(Awad et al., 2023)** are used to compare, since they are considered related to the proposed work. and the comparison shown in **Table 4** also included a deep learning-based DCHT algorithm **(Yousif et al., 2023).** It is good to mention that PESQ in general, can be used to evaluate the quality and intelligibility of the enhanced speech signal. As shown in **Table 4**, the proposed hybrid system consistently outperformed its competitors on all metrics. At a signal-to-noise ratio of 0 dB, the pink and white noises produced by the hybrid systems have scores of 2.36 and 2.29, respectively. The scores for the baseline approaches were usually between 1.0 and 1.06; our outcomes more than doubled those scores. This suggests that perceptual clarity has significantly improved. Additionally, under −5 dB conditions, where conventional models frequently degraded to −2.8 dB or lower, the hybrid system demonstrated strong SNRseg performance, with values reaching 2.25 at 0 dB SNR, in contrast to near-zero or negative values for other methods. This demonstrates how resilient the hybrid estimator is at maintaining signal energy even in the face of extreme noise. This advantage is also consolidated by the scores in Cbak, where it can be observed that our hybrid system steadily and gradually diminished nuisances at the background (e.g., 2.41 for White noise at 0 dB), while other approaches either did not report such a score or produced considerably lower values. Ultimately, the PKDMH system achieved values above 2.0 for Covl scores, which measure the overall acceptability of the improved speech. This suggests a well-balanced trade-off between quality and intelligibility, an area where other models struggled, rarely surpassing 1.08. Together, these findings demonstrate the PKDMH structure's superiority over statistical and deep learning-based baselines in demanding acoustic settings. The steady improvements in PESQ, SNRseg, Cbak, and Covl demonstrate its ability to reduce noise while maintaining perceptual quality and intelligibility at the same time.

Table 4. Comparative evaluation of different methods with the proposed method.

SNR db	Type	Methode	PESQ	SNRseg	cbak	covl
0 db SNR	white	Laplace + Laplace	1.06	0.04	–	1
		Gaussian+Gaussian	1.06	0.04	–	1
		FCNN-based DCHT	–	0.07	1.81	–
		Our hybrid estimator	**2.36**	**2.25**	**2.41**	**2.18**
	pink	Laplace + Laplace	1	0.12	–	1.08
		Gaussian+Gaussian	1	0.12	–	1.08
		FCNN-based DCHT	–	-1.03	1.63	–
		Our hybrid estimator	**2.29**	**1.40**	**2.22**	**2.15**
-5 db SNR	white	Laplace + Laplace	1.03	-2.84	–	1
		Gaussian+Gaussian	1.03	-2.68	–	1
		FCNN-based DCHT	–	-1.54	1.63	–
		Our hybrid estimator	**2**	**1.22**	**2.05**	**1.61**
	pink	Laplace + Laplace	1	-2.84	–	1.03
		Gaussian+Gaussian	1	-2.84	–	1.03
		FCNN-based DCHT	–	-2.61	1.44	–
		Our hybrid estimator	**1.9**	**0.34**	**1.84**	**1.57**

3.4 Analysis of the Improvement Rate Compared to Baseline Deep Learning Methods.

In this section, a differential evaluation was carried out in relation to unprocessed noisy signals in order to measure the extent of enhancement brought about by the suggested hybrid estimator. The differential evaluation measures the rate of improvement between the noisy and the enhanced signals. Using five objective metrics—PESQ, SNRseg, Cbak, Covl, and Csig—the comparison focused on three different stationary and non- stationary types of noise: White, Pink, and F16—under three SNR conditions (0 dB, 5 dB, and 10 dB). The net improvement attained by each method in comparison to the noisy baseline is shown in **Figs. 3 to 5**. The proposed estimator is compared to speech enhancement that uses DL to show its performance against these types of related works.

The findings unequivocally show that the suggested hybrid system produces noticeably higher gains on every evaluation metric. The hybrid estimator, for example, outperformed both CNN and FCNN methods **(Al-Zubaidi et al., 2024)**, whose improvements stayed below +0.3 for PESQ and under +0.6 for Csig, improving PESQ by +1.051, Csig by +1.226, and SNRseg by +7.005 in the case of white noise at 0 dB SNR. Patterns under pink and F16 noise were similar. The hybrid system demonstrated PESQ gains of +1.049 (white noise) and +0.903 (pink noise) at 5 dB SNR, while SNRseg increased by up to +6.1 when F16 noise was present. All approaches demonstrated some improvement at higher SNR conditions (10 dB), but the hybrid estimator consistently produced the largest differential gains. For instance, under pink noise, the hybrid approach improved PESQ by +0.836 as opposed to +0.362 for FCNN, and Covl increased to +1.108 as opposed to +0.837. These results confirm that the proposed PKDMH design achieves noticeably higher restoration from the noisy baseline than other benchmark methods, in addition to improving speech's absolute quality and intelligibility. Its robustness and effectiveness in actual noisy environments are highlighted by its superior differential performance across different types of noise and SNR levels.

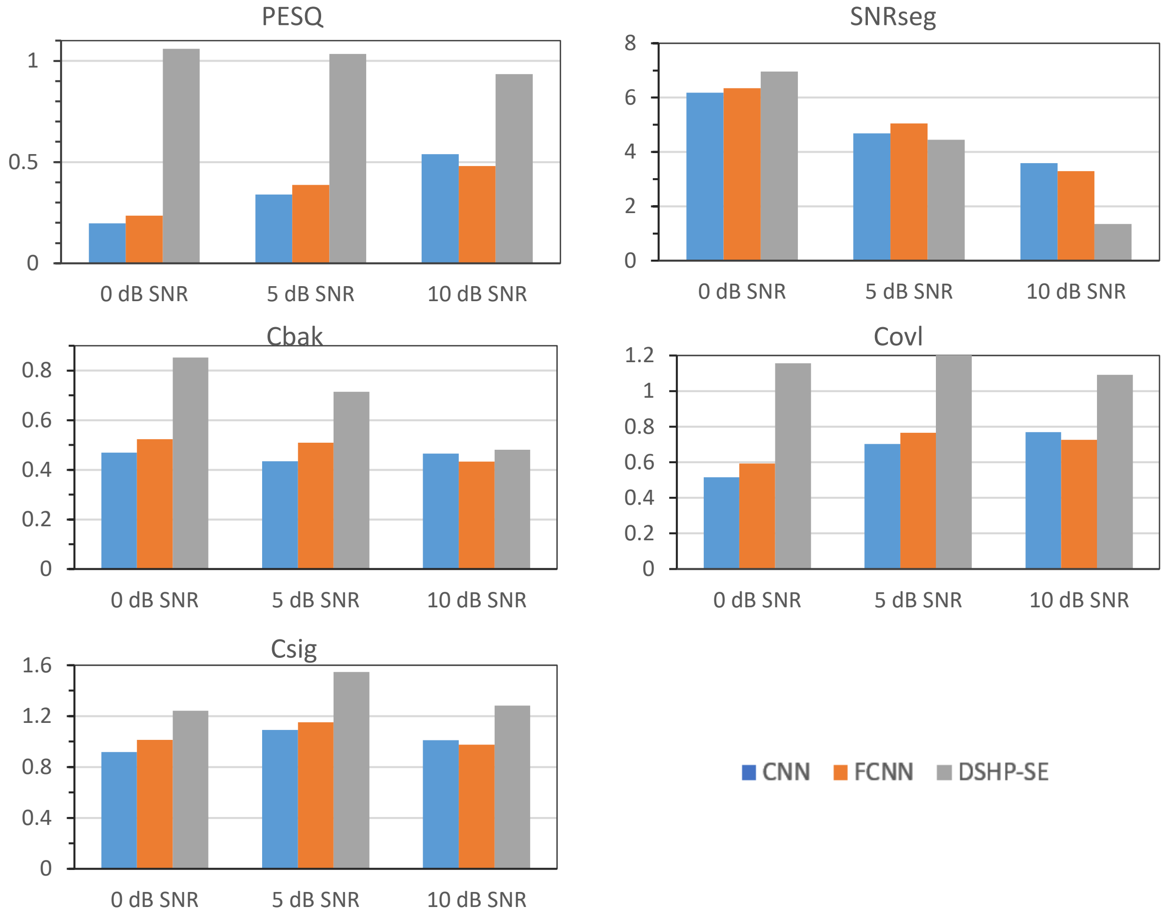
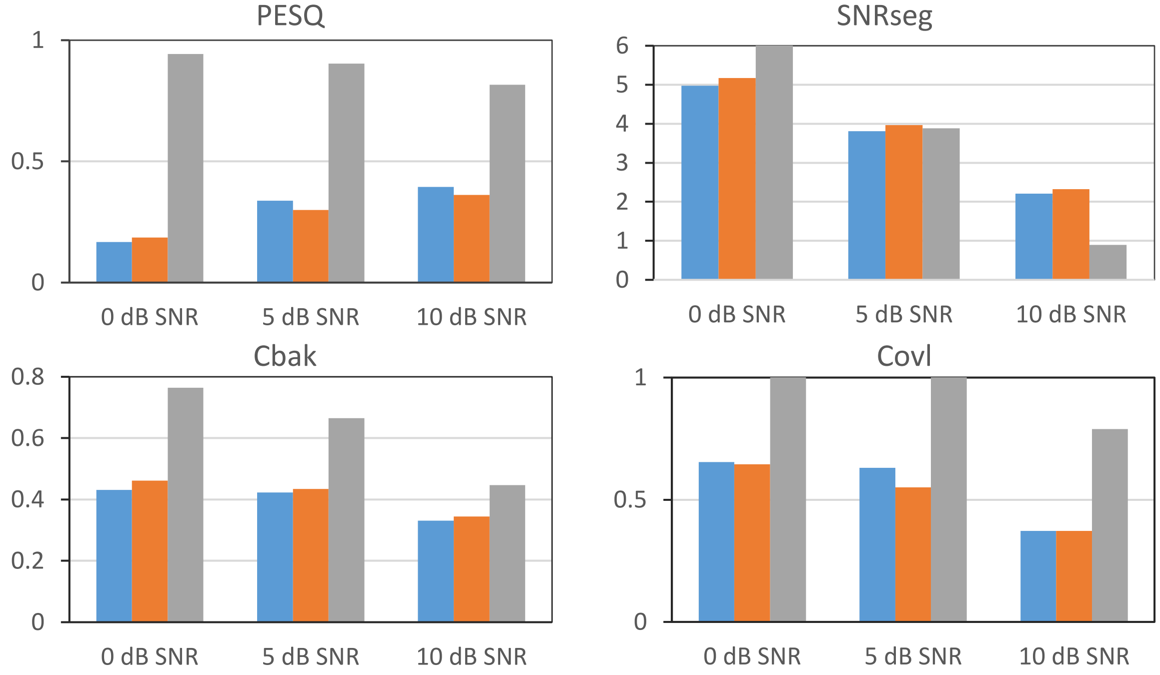


Figure 3. Comparison of different metrics across three SNR levels based on the white noise condition.



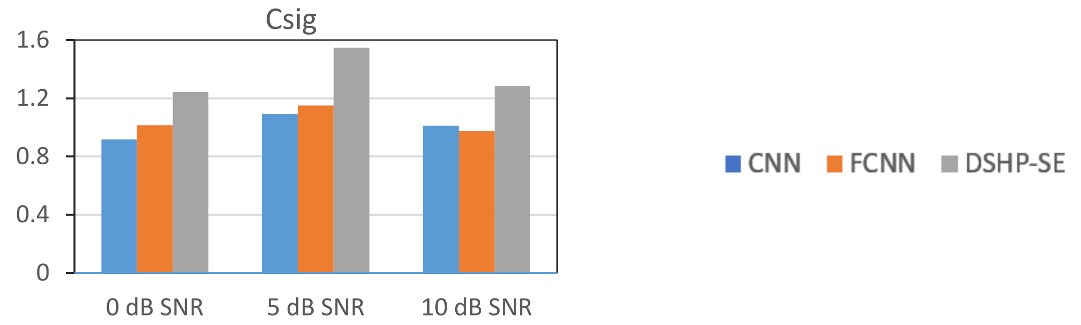


Figure 4. Comparison of different metrics across three SNR levels based on the pink noise condition.

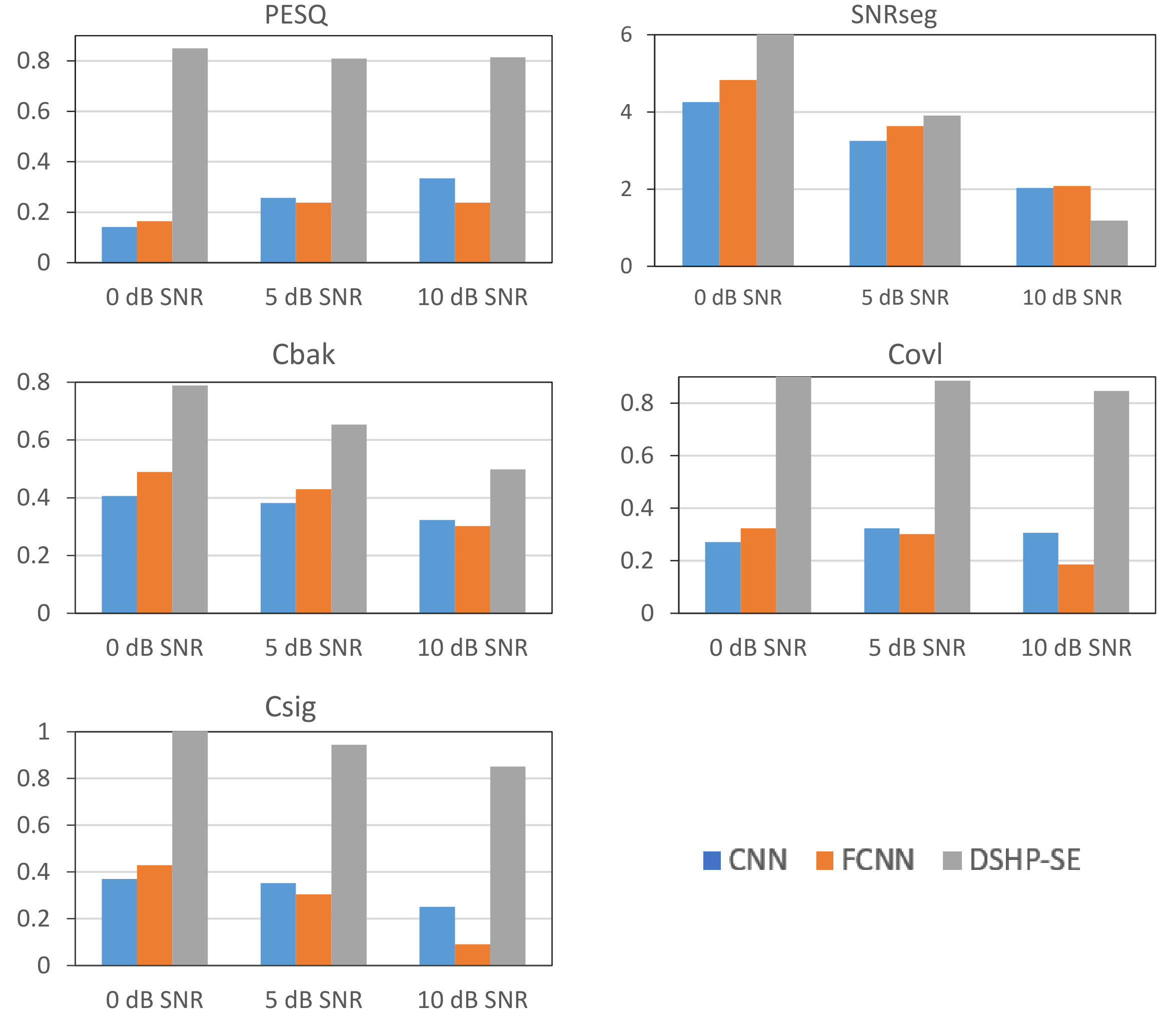


Figure 5. Comparison of different metrics across three SNR levels based on the F16 noise condition.

The performance of the PKLT→DMH cascade originates from this complementary function in the two stages and their sequential connection. In the first pass, PKLT applies a perceptually motivated subspace filtering to suppress inaudible and perceptually irrelevant noise components while retaining the dominant speech structures. Such a processing stage would yield a spatially smoother estimate size with lower white noise energy, but it may leave residual harmonic and structured noise components, especially under low-SNR and non-stationary conditions.

This second stage, DMH, specializes in detecting the remaining part of them. Using two-sided SNR estimation and harmonic-residual masking, DMH concentrates on voiced segments and harmonics that are not covered at all by subspace projection only. Used in conjunction with PKLT, DMH can process a cleaner input signal to perform harmonic regeneration and masking more accurately and with minimum speech distortion. Consequently, this cascaded order allows for progressive noise suppression; PKLT discriminates against perceptually irrelevant noise in a first stage, and then DMH operates on the result, attenuating remaining and harmonic noise artifacts.

4. CONCLUSIONS

In this work, a novel two-stage hybrid speech enhancement (SE) estimator called PKDMH is introduced, which combines the Perceptually motivated Karhunen–Loève Transform (PKLT) and the Dual-Masking Harmonic-based (DMH). The PKLT stage uses psychoacoustic masking and subspace projection to reduce perceptually unimportant noise while keeping the important speech structures in place, leading to a perceptually enhanced version of the signal being fed into the DMH stage. This stage further refines the target signal, improving voiced speech by transforming residual and harmonic noise patterns. Extensive experimentation using the TIMIT dataset, which included five noise types (White, Pink, F16, Airport, and Car) at various signal-to-noise ratios (SNRs), showcased the effectiveness of the PKDMH framework. It was evaluated against its constituent components and state-of-the-art methods using PESQ, STOI, SNRseg, Cbak, and Covl metrics. Results showed that the PKLT→DMH structure constantly achieves superior results than other competitors and even single methods, especially when SNR is low. Further comparisons were compared over its ingredients and state-of-the-art methods, including PESQ, STOI, SNRseg, Cbak, and Covl. Results showed that the PKLT→DMH structure constantly achieves superior results than other competitors and even single methods, especially when SNR is low. Finally, we note that the PKDMH framework is a strong perceptually motivated approach that is based on a sensible processing order, and future pursuit of adaptive parameter tuning or learnable modules might lead to stronger generalization in noise.

Credit Authorship Contribution Statement

Sally Taha Yousif: Data Curation, Original Draft Preparation, and Methodology. Basheera M. Mahmmod: Methodology, Manuscript review and editing,

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFRENCES

Abdulhussain, S.H., Mahmmod, B.M., Naser, M.A., Alsabah, M., and Mustafina, J., 2021. Speech enhancement algorithm based on a hybrid estimator. *IOP Conference Series: Materials Science and Engineering*, 1090(1), P. 012102. https://doi.org/10.1088/1757-899X/1090/1/012102.

Al-Zubaidi, A.S., Abduljabbar, R.B., Mahmmod, B.M., Abdulhussain, S.H., Naser, M.A., Alsabah, M., Hussain, A., and Al-Jumeily, D., 2024. Speech enhancement algorithm using deep learning and Hahn polynomials. In: *Proceedings of the 17th International Conference on Developments in eSystems Engineering (DeSE)*, pp. 42–47. IEEE. https://doi.org/10.1109/DeSE63988.2024.10911938.

Awad, H.A., Hameed, S.M., Mahmmod, B.M., Abdulhussain, S.H., and Hussain, A.J., 2023. Dual stages of speech enhancement algorithm based on super Gaussian speech models. *Journal of Engineering, University of Baghdad*, 29(9), pp. 1–13. https://doi.org/10.31026/j.eng.2023.09.01.

Boll, S.F., 1979. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27(2), pp. 113–120. https://doi.org/10.1109/TASSP.1979.1163209.

Cappe, O., 1994. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor. *IEEE Transactions on Speech and Audio Processing*, 2(2), pp. 345–349. https://doi.org/10.1109/89.279283.

Cohen, I., 2002. Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator. *IEEE Signal Processing Letters*, 9(4), pp. 113–116. https://doi.org/10.1109/97.995823.

Elert, G., 2016. The nature of sound--the physics hypertextbook. *physics.info*. Retrieved, pp. 6–20.

Ephraim, Y. and Van Trees, H.L., 1995. A signal subspace approach for speech enhancement. IEEE Transactions on Speech and Audio Processing, 3(4), pp. 251–266. http://dx.doi.org/10.1109/89.397090.

Ephraim, Y., and Malah, D., 1984. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(6), pp. 1109–1121. https://doi.org/10.1109/TASSP.1984.1164453.

Fu S.W., Yu C., Hsieh T.A., Plantinga P., Ravanelli M., Lu X., and Tsao Y., 2021. MetricGAN+: An Improved Version of MetricGAN for Speech Enhancement. *Interspeech*, pp. 201–205, https://doi.org/10.48550/arXiv.2104.03538.

Garofolo, J.S., Lamel, L.F., Fisher, W.M., Fiscus, J.G., Pallett, D.S., and Dahlgren, N.L., 1993. *TIMIT Acoustic-Phonetic Continuous Speech Corpus*. Linguistic Data Consortium, Philadelphia. https://doi.org/10.35111/17gk-bn40.

Harris, F.J., 1978. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1), pp. 51–83. https://doi.org/10.1109/PROC.1978.10837.

Hasan, T. and Hasan, M.K., 2010. MMSE estimator for speech enhancement considering the constructive and destructive interference of noise. *IET Signal Processing*, 4(1), pp. 1–11. https://doi.org/10.1049/iet-spr.2008.0114.

Hattaraki, S.M. and Kambalimath, S.G., 2024. Enhancing speech intelligibility in hearing aids using spectral subtraction. *Advanced Engineering Science*, 56(7), pp. 4793–4799.

Hu Y., Liu Y., Lv S., Xing M., Zhang S., Fu Y., Wu J., Zhang B., and Xie L., DCCRN: Deep Complex Convolution Recurrent Network for Phase-Aware Speech Enhancement. *Interspeech*, 2020, pp. 2472–2476. https://doi.org/10.21437/Interspeech.2020-2537.

Hu, Y. and Loizou, P.C., 2007. Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(1), pp. 229–238. https://doi.org/10.1109/TASL.2007.911054.

Huang, J. and Zhao, Y., 2000. A DCT-based fast signal subspace technique for robust speech recognition. IEEE Transactions on Speech and Audio Processing, 8(6), pp. 747–751. https://doi.org/10.1109/89.876314.

Jabloun, F. and Champagne, B., 2003. Incorporating the human hearing properties in the signal subspace approach for speech enhancement. *IEEE Transactions on Speech and Audio Processing*, 11(6), pp. 700–708. https://doi.org/10.1109/TSA.2003.819954.

Jerjees, S.A., Mohammed, H.J., Radeaf, H.S., Mahmmod, B.M., and Abdulhussain, S.H., 2023. Deep learning-based speech enhancement algorithm using Charlier transform. *Proceedings of the 15th International Conference on Developments in eSystems Engineering (DeSE)*, pp. 100–105. IEEE. https://doi.org/10.1109/DeSE58274.2023.10099854.

Kolbæk, M., Tan, Z.H. and Jensen, J., 2016. Speech intelligibility potential of general and specialized deep neural network-based speech enhancement systems. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(1), pp. 153–167. https://doi.org/10.1109/TASLP.2016.2628641.

Loizou, P.C., 2013. *Speech Enhancement: Theory and Practice*. 2nd ed. Boca Raton: CRC Press.

Michelsanti D., Tan Z.H., Zhang S.X., Xu Y., Yu M., Yu D., and Jensen J.,2021. An Overview of Deep-Learning-Based Audio-Visual Speech Enhancement and Separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1368–1396. https://doi.org/10.1109/TASLP.2021.3066303.

Nabi, W., Aloui, N. and Cherif, A., 2016. Speech enhancement in dual-microphone mobile phones using Kalman filter. *Applied Acoustics*, 109, pp. 1–4. https://doi.org/10.1016/j.apacoust.2016.02.009.

Nasir, R.J. and Abdulmohsin, H.A., 2025. A hybrid method for speech noise reduction using Log-MMSE. *Iraqi Journal of Science*, 66(2), pp. 860–875. https://doi.org/10.24996/ijs.2025.66.2.24.

Natarajan, S., Al-Haddad, S.A.R., Ahmad, F.A., Kamil, R., et al., 2025. Deep neural networks for speech enhancement and speech recognition: A systematic review. *Ain Shams Engineering Journal*, 16(7), Article 103405. https://doi.org/10.1016/j.asej.2025.103405.

Ochieng, P., 2023. Deep neural network techniques for monaural speech enhancement and separation: state of the art analysis. *Artificial Intelligence Review*, 56(Suppl 3), pp. 3651–3703. https://doi.org/10.48550/arXiv.2212.00369.

Plapous, C., Marro, C. and Scalart, P., 2006. Improved signal-to-noise ratio estimation for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(6), pp. 2098–2108. https://doi.org/10.1109/TASL.2006.872626.

Rix, A.W., Beerends, J.G., Hollier, M.P. and Hekstra, A.P., 2001. Perceptual evaluation of speech quality (PESQ)—A new method for speech quality assessment of telephone networks and codecs. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2, pp. 749–752. https://doi.org/10.1109/ICASSP.2001.941023.

Roy, S.K. and Paliwal, K.K., 2021. Robustness and sensitivity tuning of the Kalman filter for single-channel speech enhancement in real-life noise conditions. *Signals*, 2(3), P. 27. https://doi.org/10.3390/signals2030027.

Scalart, P. and Filho, J.V., 1996. Speech enhancement based on a priori signal-to-noise ratio estimation. In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2, pp. 629–632. https://doi.org/10.1109/ICASSP.1996.543199.

Scalart, P., Vieira Filho, J. and Chiquito, J.G., 1996. On speech enhancement algorithms based on MMSE estimation. In: *1996 8th European Signal Processing Conference (EUSIPCO 1996)*, pp. 1–4. IEEE. https://doi.org/10.5281/zenodo.36358.

Shi, S., Paliwal, K. and Busch, A., 2023. On DCT-based MMSE estimation of short-time spectral amplitude for single-channel speech enhancement. *Applied Acoustics*, 202, P. 109134. https://doi.org/10.1016/j.apacoust.2022.109134.

Soon, Y., Koh, S.N. and Yeo, C.K., 1998. Noisy speech enhancement using discrete cosine transform. *Speech Communication*, 24(3), pp. 249–257. https://doi.org/10.1016/S0167-6393(98)00019-3.

Stylianou, Y., 2001. Removing noise from speech using spectral subtraction and harmonicity-based masking. *Speech Communication*, 34(3), pp. 271–288. https://doi.org/10.1016/S0167-6393(00)00052-3.

Taal C.H., Hendriks R.C., Heusdens R., and Jensen J., 2011. An algorithm for intelligibility prediction of time–frequency weighted noisy speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7), pp. 2125–2136. https://doi.org/10.1109/TASL.2011.2114881.

Verteletskaya, E. and Simak, B., 2010. Noise reduction based on modified spectral subtraction method. *The 17th International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 233–236.

Vetter, R., 2001. Single-channel speech enhancement using MDL-based subspace approach in bark domain. *International Conference on Acoustics, Speech, and Signal Processing (ICASSP'01*), 1, pp.641–644. https://doi.org/10.1109/ICASSP.2001.940913.

Yousif, S.T., and Mahmmod, B.M., 2025. Speech enhancement algorithms: A systematic literature review. *Algorithms*, 18(5), Article 272. https://doi.org/10.3390/a18050272.

Zwicker, E. and Fastl, H., 1990. *Psychoacoustics*. Berlin: Springer-Verlag. https://doi.org/10.1007/978-3-540-68888-4.

Zwicker, E. and Fastl, H., 1999. *Psychoacoustics: Facts and Models*. 2nd ed. Berlin: Springer. https://doi.org/10.1007/978-3-662-03976-6.

مقدّر هجين جديد لتحسين إشارة الكلام

سالي طه يوسف، بشيرة محمد رضا محمود*

قسم هندسة الحاسوب، كلية الهندسة، جامعة بغداد، بغداد، العراق

الخلاصة

تقترح هذه الورقة البحثية مُقدِّرًا هجينًا لتحسين جودة الكلام، يدمج تحويل كارونين–لوف المُحفَّز إدراكيًا (PKLT) مع خوارزمية التوافقيات ذات القناع المزدوج (DMH) في إطار عمل موحد يُسمى .PKDMH تكمن الجدة الرئيسية في الجمع بين إسقاط الفضاء الفرعي الإدراكي وكبح البقايا التوافقية، مما يُمكِّن النظام من إزالة الضوضاء مع الحفاظ على الإشارات الطيفية ذات الصلة بالكلام. يقوم PKLT أولًا بإسقاط الفضاء الفرعي الإدراكي وكبح المكونات غير المسموعة، وبعد ذلك تُزيل DMH البقايا المتبقية من النطاق العريض والتوافقيات. تم تقييم نظام PKDMH المقترح باستخدام مجموعة بيانات TIMIT الملوثة بخمسة أنواع من الضوضاء – الضوضاء البيضاء، والوردية، وF16، وضوضاء المطار، وضوضاء السيارات – عبر خمسة مستويات من نسبة الإشارة إلى الضوضاء (–10 ديسيبل، –5 ديسيبل، 0 ديسيبل، +5 ديسيبل، +10 ديسيبل). استخدم التقييم الموضوعي مقاييس إدراكية ومقاييس على مستوى الإشارة، بما في ذلك PESQ و STOIو SNRsegو Csigو Cbakو .Covlأظهرت النتائج تحسينات كمية واضحة، حيث بلغ متوسط مكاسب PESQ 1.099 و0.888 و0.824 للضوضاء البيضاء والوردية وضوضاء F16 على التوالي. تُبرز هذه النتائج الميزة العملية لتقنية PKDMH المتتالية، إذ تُوفر استراتيجية تحسين أكثر موثوقية في ظل ظروف صوتية متغيرة للغاية ونسبة إشارة إلى ضوضاء منخفضة.

الكلمات المفتاحية :القناع الثنائي–التوافقDMH، وضوح الكلام، التحويل كارهوينن–لوف المدرك سمعياًPKLT،PKDMH ، معايير جودة الأداء، تحسين الكلام.