Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

*Digital Object Identifier 10.1109/ACCESS.2017.DOI*

# Speech Enhancement Algorithm Based on Super-Gaussian Modeling and Orthogonal Polynomials

**BASHEERA M. MAHMMOD[1], ABD RAHMAN RAMLI[2], THAR BAKER[3], FERAS AL-OBEIDAT[4], SADIQ H. ABDULHUSSAIN[1], AND WISSAM A. JASSIM[5] .**

[1]Department of Computer Engineering, University of Baghdad, Baghdad, 10071, Iraq. (e-mail: basheera.m@coeng.uobaghdad.edu.iq; sadiqhabeeb@coeng.uobaghdad.edu.iq)
[2]Department of Computer and Communication System Engineering, Universiti Putra Malaysia, Selangorm, 43400, Malaysia
[3]Department of Computer Science, Liverpool John Moores University, Liverpool L3 3AF, United Kingdom
[4]College of Technological Innovation, Zayed University, Abu Dhabi 144534, United Arab Emirates
[5]ADAPT Center, School of Engineering, Trinity College Dublin, University of Dublin, Dublin 2, Ireland

Corresponding author: Thar Baker (email: t.m.shamsa@ljmu.ac.uk; t.baker@ljmu.ac.uk), and Sadiq H. Abdulhussain (e-mail: sadiqh76@yahoo.com; sadiqhabeeb@coeng.uobaghdad.edu.iq).

**ABSTRACT** Different types of noise from the surrounding always interfere with speech and produce annoying signals for the human auditory system. To exchange speech information in a noisy environment, speech quality and intelligibility must be maintained, which is a challenging task. In most speech enhancement algorithms, the speech signal is characterized by Gaussian or super-Gaussian models, and noise is characterized by a Gaussian prior. However, these assumptions do not always hold in real-life situations, thereby negatively affecting the estimation, and eventually, the performance of the enhancement algorithm. Accordingly, this paper focuses on deriving an optimum low-distortion estimator with models that fit well with speech and noise data signals. This estimator provides minimum levels of speech distortion and residual noise with additional improvements in speech perceptual aspects via four key steps. First, a recent transform based on an orthogonal polynomial is used to transform the observation signal into a transform domain. Second, noise classification based on feature extraction is adopted to find accurate and mutable models for noise signals. Third, two stages of nonlinear and linear estimators based on the minimum mean square error (MMSE) and new models for speech and noise are derived to estimate a clean speech signal. Finally, the estimated speech signal in the time domain is determined by considering the inverse of the orthogonal transform. The results show that the average classification accuracy of the proposed approach is 99.43%. In addition, the proposed algorithm significantly outperforms existing speech estimators in terms of quality and intelligibility measures.

**INDEX TERMS** MMSE estimator, orthogonal polynomials, Speech Enhancement, Super-Gaussian distribution.

## I. INTRODUCTION

Speech is the primary means of interaction among human beings. It plays a key role in the recent communication technological era. Speech signals experience several difficult scenarios during transmission, such as interference, reverberation, and additive environmental noise. Additive noise is considered the most influential and most widespread type of noise in a real environment; therefore, Speech Enhancement Algorithms (SEAs) have been developed to deal with noisy signals, restore clean speech signals, improve speech quality and intelligibility, solve the noise pollution problem, and reduce listener fatigue [1], [2]. The process of removing noise without distorting the original speech signal is a challenging task [3]. SEAs are commonly implemented in different applications [3]–[7].

Several studies have categorized SEAs into two main groups: supervised and unsupervised methods [8]–[10]. Other works have divided SEAs into three main classes based on the techniques used to process information: spectral-subtractive algorithms [11]; algorithms based on statistical models and optimization criteria, such as Wiener filtering (WF) [1], [6] and minimum mean square error (MMSE)

algorithms [12], [13]; and sub-space algorithms [14], [15]. Another mode of classification depends on the processing domain, namely, time domain [16]–[18] and transform domain [6], [19]. Algorithms that belong to the transform domain generally compress substantial information in a signal into specific coefficients; therefore, high energy compaction capability and good spectral resolution are achieved [20], which leads to an effective noise removal process [21], [22]. The most well-known discrete transforms in the speech enhancement field are discrete Fourier transform (DFT) [12], discrete cosine transform (DCT) [21], [23], [24], discrete Krawtchouk transform (DKT) [6], discrete Tchebichef transform (DTT) [6], wavelet transform (WT) [25], and discrete Krawtchouk–Tchebichef transform (DKTT) [26]. Generally, most of the mentioned works focused on processing the magnitude of the speech signal to enhance the speech signal, however, there are other researches work on phase processing for speech enhancement [27], [28].

The probability density function (PDF) of speech and noise signals is considered a crucial point in designing a statistical speech estimator. Most conventional SEAs adopt Gaussian [3], [12], [29], Laplacian [4], [13], [30], or Gamma [31] priors to model speech signals, whereas noise is predominantly modeled as a Gaussian random process [3]. The fundamental work can be traced back to the introduction of the short-time spectral amplitude (STSA) estimator for clean speech signals by Ephraim and Malah [12]. This estimator is based on modeling speech and noise Fourier expansion coefficients as statistically independent, zero-mean, and Gaussian random variables. It is derived by minimizing the conditional mean squared error (MSE) [8]. Ephraim and Malah extended their work in [32] by using log spectral amplitude (LSA) to improve agreement with the mechanism of human hearing [23]. This estimator is efficient in reducing the musical noise (MN) phenomenon [33]. A modified LSA was proposed by Cohen [34] by modifying the gain function of the LSA estimator based on a binary hypothesis model. A combination of MMSE estimators and spectral subtraction filter was developed in [35]. Different studies have used real transforms, such as DCT [21], [24], DKT and DTT [6], and WT [36], [37], for enhancing noisy signals. These transforms are effective in noise reduction [21], [22]. The attenuation filter is not always suitable for noise interferences, and thus, Soon and Koh [29] proposed an innovative approach that minimizes the distortion of reconstructed signals by considering two cases of additive noise. This approach called the low distortion approach. It minimizes underlying speech distortion during speech enhancement process since it identifies whether the background noise is destructive or constructive for a specific sequence. That means the attenuation filter is used to reverse the process of additive noise; however, the resultant magnitude of the addition of two complex signals (speech and noise) may not always be greater than the original amplitude of speech. Therefore, using an attenuation filter leads to high distortion in speech signal [29]. Two filters, i.e., the multiplicative dual-gain Wiener filter (DGW) and subtractive filters are used in this approach. Real transforms based on an orthogonal polynomial (OP) were first used by Jassim et al. [6] to enhance noisy signals based on the WF approach in the DKT and DTT domains. If speech and noise are modeled as Gaussian priors in the real transform, then the resulting spectral gain becomes a WF, as proven by Wolfe and Godsil [8], [38].

Many SEAs have adopted super-Gaussian functions to model speech signals [31], [39] because super-Gaussian distributions have longer tails and spikier peaks, and thus, are more appropriate to represent speech signals. Moreover, a Gaussian assumption is asymptotically valid only when the size of the duration frame is longer than the span correlation of the signal under consideration [4], [39], [40]. This assumption may hold for noise components but not for speech components, which are typically estimated using relatively short (20–30 ms) duration windows [3], [4]. Different SEAs have reinforced this concept [40], [41]. In [40], the capability of Laplacian random variables to describe speech samples during voice activity intervals was proven. The selection of an appropriate PDF is based on a comparison between a speech coefficient histogram obtained from a large dataset and a non-Gaussian distribution [31]. Many researchers have adopted Laplacian or gamma PDF in their works, such as [4], [13], [30], [31], [39], [42], [43]. Although SEA performance is improved, the optimal points of speech quality and intelligibility have not been achieved because leakage occurs in speech and noise modeling. Most studies do not state the different properties of various types of noise [44]. In a single-microphone setting, improving quality and intelligibility attributes is a popular research topic [45].

Conventional SEAs require noise estimation algorithms to perform correctly [46]. Most of these algorithms suffer from residual noise and speech distortion because the details of speech signals are essentially destroyed under low signal-to-noise ratio (SNR), in addition to the difficulty of processing non-stationary noise [47]. Various SEAs have attempted to address these drawbacks, but their success depends on noise type [46]. Therefore, recent studies that utilize the noise classification process are recommended [37], [44]–[46], [48]. Noise classification is first performed, followed by SEA, which uses optimal parameters based on the selected noise type. However, no method uses noise classification to find the best noise model, which is a significant point in statistical SEAs. Accordingly, the current study proposes novel linear and nonlinear low-distortion estimators that account for constructive and destructive events based on new composite super-Gaussian representations of speech and noise signals. The new model for speech DKTT coefficients is a composite of Laplacian and gamma distributions, whereas the noise DKTT coefficient model is represented by a dual Laplacian prior. In this paper, a new estimator is proposed to avoid high distortion in speech signals in low SNR regions, minimize residual noise (including MN), and concurrently improve quality and intelligibility perceptual aspects. Accordingly, this paper focuses on deriving an optimum low-

TABLE 1: Table of Notions

| | |
|---|---|
| $\propto$ | smoothing parameter |
| $DCP$ | distribution controlling parameter |
| $p_E$ | expectation parameter |
| $x(n)$ | discrete time speech signal |
| $d(n)$ | discrete time uncorrelated noise |
| $y(n)$ | discrete time noisy signal |
| $X_l(k)$ | $k$th coefficient of speech signal |
| $D_l(k)$ | $k$th coefficient of noise signal |
| $Y_l(k)$ | $k$th coefficient of noisy signal |
| $f(x)$ | single dimension signal in the time domain |
| $F(k)$ | single dimension signal in the transform domain |
| $R_m(x)$ | $m$th order polynomial of the Krawtchouk-Tchebichef transform |
| $t_i(x)$ | $i$th weighted and normalized Tchebichef polynomial |
| $k_i(x)$ | $i$th weighted and normalized Krawtchouk polynomial |
| ${}_pF_q$ | Hypergeometric function |
| $E\{\cdot\}$ | expectation operator |
| $e_k$ | mean square error |
| $E_+$ | speech and noise are constructive event |
| $E_-$ | speech and noise are destructive event |
| $P(a,b)$ | joint statistics of $a$ and $b$ |
| $p(\cdot)$ | probability density function |
| $p$ | controlling parameter of Krawtchouk polynomial |
| $P_y(l,\cdot)$ | power for each frame |
| $\zeta_k$ | prior SNR |
| $\gamma_k$ | posterior SNR |
| $G_c^{LBSE}$ | gain of the LBSE constructive events |
| $G_d^{LBSE}$ | gain of the LBSE destructive events |
| $G_c^{NBSE}$ | gain of the NBSE constructive events |
| $G_d^{NBSE}$ | gain of the NBSE destructive events |
| $\delta_e$ | percentage of MSE improvement |

TABLE 2: Table of abbreviations

| | |
|---|---|
| DCP | distribution controlling parameter |
| DCT | discrete cosine transform |
| DFT | discrete Fourier transform |
| DGW | dual-gain Wiener filter |
| DKT | discrete Krawtchouk transform |
| DKTT | discrete Krawtchouk-Tchebichef transform |
| DMMSE | dual MMSE estimator |
| DTT | discrete Tchebichef transform |
| LBSE | linear bilateral super-Gaussian estimator |
| LGMDGW | Laplacian-Gaussian mixture-based dual-gain Wiener filter |
| LSA | log spectral amplitude |
| MMSE | minimum mean square error |
| MN | musical noise |
| MSE | mean squared error |
| NBSE | nonlinear bilateral super-Gaussian estimator |
| OP | orthogonal polynomial |
| PDF | probability density function |
| SEA | Speech Enhancement Algorithm |
| SNR | signal-to-noise ratio |
| STSA | short-time spectral amplitude |
| SVM | support vector machines |
| TSDKTE | rwo-stage-based DKT estimator |
| TSDTTE | two-stage-based DTT estimator |
| WF | Wiener filtering |
| WT | wavelet transform |

distortion estimator with models that fit well with speech and noise data signals to provide minimum levels of speech distortion and residual noise with additional improvements in speech perceptual aspects. The proposed SEA combines the advantages of Laplacian and gamma priors for modeling speech and noise signals in a real transform to provide good enhancement performance.

The rest of this paper is organized as follows. Section II describes the strategy stages of the proposed SEA and the basic mathematical aspects of DKTT and the noise classification method. The derivation of the proposed linear and nonlinear estimators is also provided in this section. Section 3 presents the evaluation of the noise classifier and the proposed estimator through a substantial comparison with several existing algorithms. Lastly, the conclusion is discussed in Section 4.

## II. THE PROPOSED SEA

The proposed SEA and its specific stages, which embed the fulfillment requirements of enhancing noisy signals, are presented in the following subsections. For more elucidation, TABLE 1 list the notions used. In addition TABLE 2 list the abbreviation used in this paper.

### A. STAGES OF THE PROPOSED SEA STRATEGY

The design of the proposed SEA is divided into five main phases. The first phase converts noisy speech into the uncorrelated domain using real transform DKTT, which is based on OP. Second, a noise classification algorithm is adopted to classify the statistical properties of noise. Then, three different sets of parameters are determined properly: the distribution controlling parameter (DCP), the expectation parameter ($P_E$), and the smoothing parameter ($\propto$). The third phase is the nonlinear bilateral super-Gaussian estimator (NBSE). The fourth phase is the linear bilateral super-Gaussian estimator (LBSE). NBSE and LBSE are two-stage estimators based on MMSE sense. They are combined in a cascading form to formulate the NLBSE. Finally, the inverse of DKTT, and then an overlap–add technique, are applied to synthesize the original speech signal back to the time domain. The proposed SEA phases are shown in FIGURE 1 and explained in the succeeding subsections.

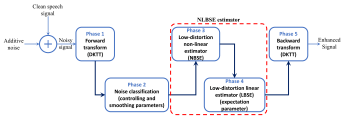



FIGURE 1: The General scheme of the proposed SEA.

### B. BASIC MATHEMATICAL ASPECTS OF DKTT

DKTT exhibits the following distinctive properties: high energy compaction, good localization [49], [50], and excellent noise suppression performance. These capabilities significantly affect the enhancement process [23], where noise can be suppressed without substantial loss of the original signal information. Moreover, real transform reduces computational complexity in noisy signal analysis and clean signal synthesis. Initially, the definition of the additive noisy signal model is expressed as follows: let $x(n)$ be the discrete time speech signal that is degraded by the uncorrelated background noise

$d(n)$ (includes white noise and color noise), which results in the following noisy signal:

$$y(n) = x(n) + d(n) \tag{1}$$

Then, y (n) is transformed into the DKTT domain to obtain $X_l(k)$ , $Y_l(k)$ and $D_l(k)$ in the $k$th transform coefficients of speech, noisy, and noise signals, respectively.

$$Y_l(k) = X_l(k) + D_l(k) \tag{2}$$

where $l$ represents the frame number. Meanwhile, the DKTT formula of the $m$th order Krawtchouk-Tchebichef transform, $R_m(x)$, which is used to transform $y(n)$ into $Y_l(k)$, is

$$R_m(x) = \sum_{i=0}^{N-1} k_i(m; p, N-1)\, t_i(x) \tag{3}$$
$$m, x = 0,\ 1,\ \ldots,\ N-1,\ N > 0,\ p \in (0, 1)$$

where $t_i(x)$ is the weighted and normalized form of the Tchebichef polynomial [51]:

$$t_i(x) = \frac{(1-N)_i\, {}_3F_2(-i, -x, 1+i; 1, 1-N; 1)}{\sqrt{(2i)!\binom{N+i}{2i+1}}} \tag{4}$$
$$i, x = 0, 1, \ldots, N-1;\ N > 0$$

where $\binom{a}{b}$ is the binomial coefficients = $\frac{a!}{b!(a-b)!}$, and $(a)_k$ represents Pochhammer symbol [52], [53].

$$(a)_k = a(a+1)(a+2), \ldots, (a+k-1) = \frac{\Gamma(a+k)}{\Gamma(a)} \tag{5}$$

Meanwhile, $k_i(m; p, N-1)$ is the weighted KP [54]:

$$k_i(m; p, N-1) = \sqrt{\frac{\binom{N-1}{m} p^m (1-p)^N - 1 - m}{(-1)^i \left(\frac{1-p}{p}\right)^i \left(\frac{i!}{(-N+1)_i}\right)}} \times {}_2F_1(-i, -m, -N+1; \frac{1}{p}) \tag{6}$$
$$i, m = 0, 1, 2, \ldots, N-1, N > 0, p \in (0, 1)$$

where ${}_3F_2$ and ${}_2F_1$ are the hypergeometric functions [55], $N$ represent the frame size, and $p$ is the controlling parameter of KP. $R_m(x)$ is used to transform the noisy signal $y(n)$ into the DKTT domain and obtain $Y_l(k)$. To transform a signal $f(x)$ from time domain to transform domain $F(k)$, the following expression is used [56]:

$$F(k) = \sum_{x=0}^{N-1} R_k(x)\, f(x) \tag{7}$$
$$k = 0, 1, \ldots, N-1$$

and to reconstruct the signal from the transform domain $F(k)$ to time domain $f(x)$, the following formula is used:

$$f(x) = \sum_{k=0}^{N-1} R_k(x)\, F(k) \tag{8}$$
$$x = 0, 1, \ldots, N-1$$

In addition, the matrix multiplication of equations (7) and (8) are as follows:

$$\mathbf{F} = \mathbf{R} \times \mathbf{f} \tag{9}$$
$$\mathbf{f} = \mathbf{R}^T \times \mathbf{F} \tag{10}$$

where $\mathbf{F}$, $\mathbf{f}$, and $\mathbf{R}$ are the matrix form of $F(k)$, $f(x)$, and $R_k(x)$, respectively, and $(\cdot)^T$ represent the matrix transpose operator. It is noteworthy that the transform domain coefficients (moments) can be used as a shape descriptor for different types of signals [57]. In addition, basis functions of OPs can be used as an approximate solution for differential equations [75].

### C. CONCEPTS OF NOISE CLASSIFICATION ALGORITHM

In order to make the proposed SEA suitable for different noise environments, a noise classification method is introduced. This method is used to find accurate models for noise signals by controlling their statistical characteristics. This process makes the PDF of the input noise signal matching the assumed distribution. Therefore, the suppression of noise will be optimized. The types of noise are classified using support vector machines (SVM) through feature extraction process. The models of SVM are trained based on eleven background noises. SVM is a very useful and popular machine learning technique for data classification [45]. SVM works well with different feature sets [58], and derived from statistical learning theorem [44]. New significant parameters are determined as stated in Section II-A based on noise classification. These parameters are defined in related sections.

#### 1) Features extraction

There are two sets of features used in this work; the mean of normalized power and the mean of the standard deviation. Features are extracted based on the normalized sub-band noise. Note that, the number of partitions of the sub-band power is 25 with length equal to 16 samples, which are experimentally enough. There are 50 features calculated to realize the corresponding noise classification model. According to the noise type, the corresponding DCP are selected. Specifically, DCP control the amplitude and standard deviation of the assumed noise PDF. The power for each frame is calculated as:

$$P_y(l, k) = [Y^2(l, 1), Y^2(l, 2), \ldots, Y^2(l, N)]^T \tag{11}$$

The normalized power feature can be obtained as follows:

$$P_{y-norm}(l,k) = [Y^2_{norm}(l,1), Y^2_{norm}(l,2), \ldots, Y^2_{norm}(l,N)]^T \tag{12}$$

where $Y^2_{norm}(l,k)$ is the normalized power in the $k$th moment, and its formula is:

$$Y^2_{norm}(l,k) = \frac{Y^2(l,k)}{\sum_{k=1}^{N} Y^2(l,k)} \tag{13}$$

From the normalized power and for each sub-band, the mean power and the standard deviation are calculated. To find the mean power, the length of each sub-band ($L$) is calculated first as:

$$L = \frac{N}{J} \tag{14}$$

where, $J$ is the total number of sub-bands. Then, the average power will be:

$$\mathcal{P}_{j,l} = \frac{\sum_{s}^{e} P_{n,l}}{L} \tag{15}$$

where $j$ is the sub-band number for each frame. $s = (j-1) \times (L+1)$, represents the index of starting sample. $e = j \times L$, represents the index of ending sample.

The first feature, the mean, is:

$$\mu_j = \frac{1}{N_{LF}} \sum_{i=1}^{N_{LF}} \mathcal{P}_{j,i} \tag{16}$$

where , $N_{LF}$ is the number of initial frames. The second feature, the standard deviation, is:

$$S_j = \sqrt{\frac{1}{N_{LF}-1} \sum_{i=1}^{N_{LF}} \left|\mathcal{P}_{|,\rangle} - \mu_j\right|^2} \tag{17}$$

Then feature vector is constructed based on these features (mean (16) and standard deviation (17)) using concatenation.

#### 2) Training of SVM Model

SVM classifier is implemented to determine the type of noise from the six initial frames of the speech signal. SVM designed for binary classification problem to solve multi-class classification problem. In this work, "one- against-one" approach is performed, which is faster to train and seems preferable for problems with a large number of classes [59] and it is based on voting strategy. For a problem with $C$ classes, the total number of classifiers will be $c(c-1)/2$, and each of them trains data from two classes [44]. Therefore, in this work, there are 55 classifier. The six initial frames from each speech segment are used for feature extraction to calculate feature vectors through performing DKTT on the windowed noisy speech. 400 speech files are taken. 100 files for training phase and 300 files for testing phase. The speech signals are corrupted by eleven types of noise, which are considered the most dominate noise in the environment. The length of training and testing data for each level of SNR is about 25 *ms* to get a stationary segment of speech signal. 5500 segments of noisy speech signal are used as training set. These numbers of files comes from 11 types of noise, 5 levels of SNRs, and 100 speech files that are used for training phase. DKTT is used with p=0.5 to provide an appropriate localization and symmetry properties that facilities the mathematical calculations.

#### 3) Testing of SVM Model

For testing phase, 300 clean speech files are chosen from TIMIT dataset [44]. The speech files denoted by 'SA1' and 'SA2' for males and females speakers. Eleven types of noise are used in testing phase with the five levels of SNR. Therefore, there are totally 16500 files for testing phase. Each noise has different set of features that distinguish between noise types. The noise is judge during the initial six frames of the noisy speech signal, which are considered noise only frames. Then the noise classification is carried out based on these features. In this work, "one- against-one" approach is performed. This approach involves constructing a classifier for each pair of classes resulting in multi classifiers. And it is based on voting strategy to combine the 55 classifiers. For the test point, each binary classifier gives one vote for the winning class and the point is labeled with the class having most votes.

For more explanation about classification method, let m and n denote two classes chosen out of the given noise types, then the training data for class pair mn that corresponding class labels z can be expressed as follows [44]:

$$D^{mn} = \{(r_i, z_i)|_{\mathbf{r}_i} \in \mathbb{R}, Z_i \in \{-1,1\}\}_{i=1}^{2M} \tag{18}$$

where, $M$ is the number of initial frames that are equal to six. The decision function for noise class pair mn is defined by:

$$f_{mn}(\mathbf{r}) = \sum_{\mathbf{r}_i \in sv} \alpha_i^{mn} z_i K(\mathbf{r_i}, \mathbf{r}) + b^{mn} \tag{19}$$

where, $\alpha_i^{mn}$ is from the solution of the quadratic programming problem, $b^{mn}$ represents the optimized bias, and $K$ denotes the Kernel function. As mentioned, voting strategy is applied for each binary classifier gives one vote for its winner class, and feature vector $\mathbf{r}$ is designated to be in a class with the most votes. The noise type of the $l$th frame corresponding to $\mathbf{r}$ is given by:

$$C_{frame} = \underset{m=1,\ldots,11}{\operatorname{argmax}} \sum_{n \neq m, n=1}^{11} sgn(f_{mn}(\mathbf{r})) \tag{20}$$

### D. PROPOSED MMSE ESTIMATORS

Linear and nonlinear estimators are proposed in this paper. These estimators are based on statistical approaches, where an enhanced signal is obtained mathematically by optimizing the dissolvable error criteria (MSE). The linear estimator (LBSE) is based on the linear WF notion because a linear relationship exists between the observed data and the estimated

signal. Meanwhile, the nonlinear MMSE estimator (NBSE) is based on the statistical analysis notion, which requires knowledge regarding speech and noise probability distributions [3]. The analytical solution for the proposed estimators is derived in this section. Each of these estimators has two gains, and each gain deals with a constructive or destructive event. Thus, each estimator is considered a bilateral gain.

### 1) Proposed Non-Linear Bilateral Super-Gaussian Estimator (NBSE)

In this estimator, the models for speech and noise transform coefficients are assumed to be statically independent super-Gaussian random variables. The main objective is to find a nonlinear estimate of the interest factors (clean signal) based on a given set of parameters (noisy signal). In the NBSE estimator, the statistical model for speech DKTT components is assumed to be a composite distribution of Laplacian and gamma PDFs (please see (28)). Meanwhile, dual Laplacian distribution is used to model noise signal (please see (32)). Where the dual Laplacian distribution corresponds to two Lapalcian PDFs with different parameters have been combined to achieve the new distribution. The probability distribution of speech is exhibited in FIGURE 3a. In this paper, eleven types of noise are used. White noise is presented in FIGURE 3b. FIGURE 3 shows that better fitting is obtained for the assumed speech and noise DKTT PDFs than for the other presented density functions.

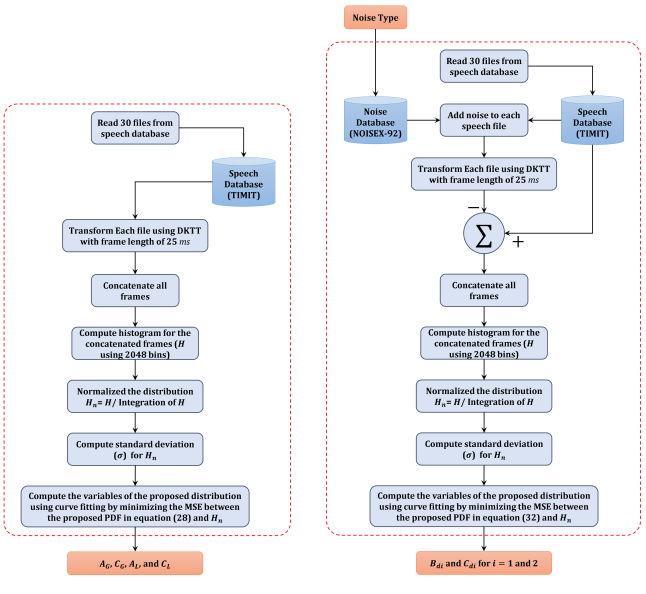
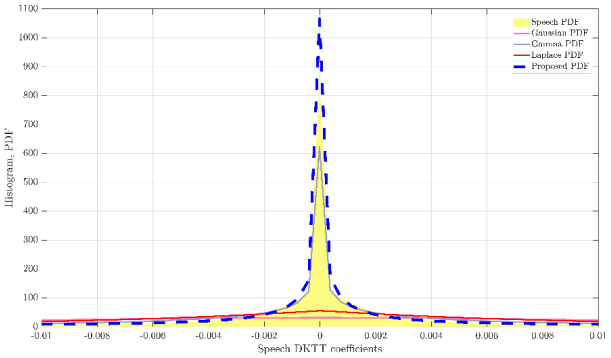
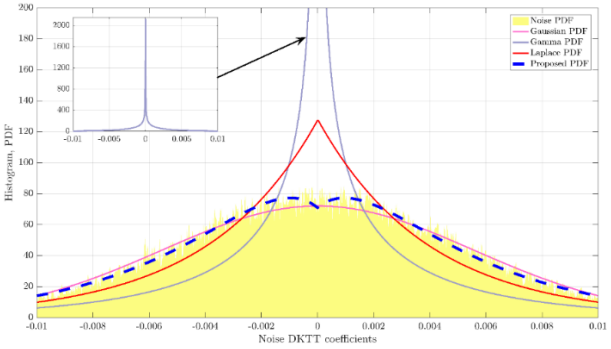



FIGURE 2: Step used to find the DCPs for the fitting model of: (a) speech signal, and (b) noise signal

Evidently, the assumed composite PDF is more accurate and provide better fitting with the DKTT data than the Gaussian, Laplacian, and gamma distributions. The enlarged section shows that the gamma prior has an extremely high value, making it inappropriate for representing DKTT data,




(a)




(b)

FIGURE 3: The proposed pdf of (a) clean speech (b) white noise DKTT coefficients verses other pdfs.

because it diverges when the argument approaches zero. FIGURE 2 shows the procedure used to find the fitting parameters for clean and noise signals.

The change in the external appearance of the proposed PDFs is controlled by DCP. Thus, noise reduction can be realized without significant loss in intelligibility. In general, significant noise reduction leads to serious degradation in speech intelligibility [60]. The speech signal model has four DCPs. One is for the gamma prior, i.e., $A_G = 0.7604$ , which controls the gamma PDF amplitude, and one is for $C_G = 1$, which controls the standard deviation. The other two parameters control the Laplacian amplitude and standard deviation, which are $A_L = 0.1839$ and $C_L = 0.03$, respectively. Meanwhile, the distribution of the DKTT noise coefficients has 44 different DCP values because the eleven types of noise have four DCPs each. $B_{d1}$ and $B_{d2}$ control the amplitude value of the dual Laplacian PDF. $C_{d1}$ and $C_{d2}$ control the standard deviation value of the dual Laplacian PDF. The second parameter found based on noise classification is $\propto$. It is a significant factor in the decision-directed approach, where the former is used to estimate a priori SNR [12]. Ideally, $\propto$ must be small during the transient parts of speech to respond faster to sudden changes in speech signals, whereas it must be large during the steady-state segments of speech to control the level of MN [3]. The optimum values of DCP and $\propto$ are listed in TABLE 3 according to noise type.

TABLE 3: DCP and $\propto$ for different types of noise

| Noise Type | $B_{d1}$ | $B_{d2}$ | $C_{d1}$ | $C_{d2}$ | $\propto$ |
|---|---|---|---|---|---|
| White | 1.47 | -0.4349 | 1 | 0.4749 | 0.94 |
| Babble | 0.6292 | 0.3497 | 1 | 0.1074 | 0.96 |
| F16 | 0.7685 | 0.2114 | 1 | 0.4965 | 0.95 |
| Pink | 1.073 | -0.0435 | 1 | 0.1342 | 0.98 |
| Speech Shaped | 0.3739 | 0.564 | 1 | 0.1762 | 0.95 |
| Buccaneer Jet | 2 | -1 | 1 | 1 | 0.94 |
| Destroyer Engine | 0.7081 | 0.2418 | 1 | 0.2086 | 0.97 |
| Destroyer operations | 0.2188 | 0.7338 | 1 | 0.5643 | 0.97 |
| Leopard | 0.3984 | 0.5927 | 1 | 0.0373 | 0.87 |
| M109 | 0.3292 | 0.6015 | 1 | 0.2527 | 0.95 |
| Factory | 0.5439 | 0.437 | 1 | 0.1047 | 0.95 |

FIGURE 4 shows the PDF distribution for the other types of noise which confirm the accurate mapping of the proposed model.

The objective of the proposed NBSE is to find $\hat{X}_k$ by minimizing the MSE between $\hat{X}_k$ and $X_k$. NBSE and LBSE have two gains each, namely, attenuation and amplification, based on the low distortion approach. The derivation begins with the MSE formula:

$$e_k = E\left\{\left(X_k - \hat{X}_k\right)^2\right\} \tag{21}$$

where $e_k$ indicates the MSE and $E\{.\}$ signifies the expectation operators. The analytical solution for NBSE and its gain functions are explained through the computation steps below. The two conditions that summarize the two mutually exclusive events must be defined first [22], [29] as follows:
$E_+$: speech and noise are constructive when $X_l(k)\, D_l(k) \geq 0$
$E_-$: speech and noise are destructive when $X_l(k)\, D_l(k) < 0$

The additive noisy signal model is expressed in (1). Then, the observed signal y(n) is transformed into the DKTT domain as indicated in (2). In NBSE, no linear relation exists between $\hat{X}_k$ and $Y_k$. Therefore, the formula for MSE in (7) must be minimized by resolving the expected value. For readability, the moment index is written as a subscript and the frame index is omitted because the work is an up-to-date frame. The expectation formula can be expressed as

$$e_k = \int_0^\infty \int_0^\infty \left(X_k - \hat{X}_k\right)^2 p\left(X_k, Y_k\right) dX_k dY_k \tag{22}$$

where $P\left(X_k, Y_k\right)$ is the joint statistics of $X_k$ and $Y_k$. Thereafter, the symbol $(\hat{\cdot})$ will refer to the estimation operation. Joint probability is converted into conditional probability based on conditional probability theory, as follows [60]:

$$e_k = \int_0^\infty p\left(Y_k\right) \int_0^\infty \left(X_k - \hat{X}_k\right)^2 p(X_k|Y_k) dX_k dY_k \tag{23}$$

To minimize MSE, the inner integral in Equation (9) must be minimized for the observation vector [61] by taking its derivative with respect to $\hat{X}_k$ and its equality to zero:

$$\hat{X}_k = \int_{-\infty}^{\infty} X_k p\left(X_k|Y\right) dx = E(X_k|Y_k) \tag{24}$$

The general definition of the conditional expectation is based on conditional probability, as follows:

$$E\left[X_k|Y_k\right] = \int_{-\infty}^{\infty} x_k p\left(x_k|Y_k,\right) dx_k \tag{25}$$

which can be solved using joint and merging probabilities, as follows:

$$E\left[X_k|Y_k\right] = \frac{\int_{-\infty}^{\infty} x_k p\left(x_k, Y_k\right)}{\int_{-\infty}^{\infty} p\left(x_k, Y_k\right)} dx_k \tag{26}$$

Therefore, a priori knowledge regarding the PDFs of speech and noise coefficient distributions is necessary. Basically, the final NBSE output to obtain the estimated signal is

$$\hat{x}_k^{\text{NBSE}} = f_k E\left[X_k|Y_k, E_+\right] + \left(1 - f_k\right) E\left[X_k|Y_k, E_-\right] \tag{27}$$

The polarity estimator parameter, which is denoted as $f_k$, controls the event probability of each condition [22], [29]. $f_k$ is assumed to be ideal in this work. The modeling of a speech signal is defined as ($F_{\text{Speech}}$) and assumed to be a composite of the gamma and Laplacian priors, as follows:

$$F_{Speech} = A_G F_G\left(x_k, C_G\, \sigma_{Gx_k}\right) + A_L F_L\left(x_k, C_L b_{Lx_k}\right) \tag{28}$$

The definition of gamma density in the proposed work is given by

$$F_G\left(x_k, \sigma_{x_k}\right) = \frac{\beta_k^\alpha\, x_k^{\alpha-1}\, e^{-\beta_k x_k}}{\Gamma(\alpha)} \tag{29}$$

For readability, $\sigma_{x_k} = C_G\, \sigma_{Gx_k}$ and the variance of the gamma PDF is $\sigma_{x_k}^2 = \frac{\alpha}{\beta^2}$. When $\alpha = 0.5$ and $\Gamma(0.5) = \sqrt{\pi}$ are considered, the resulting gamma function is

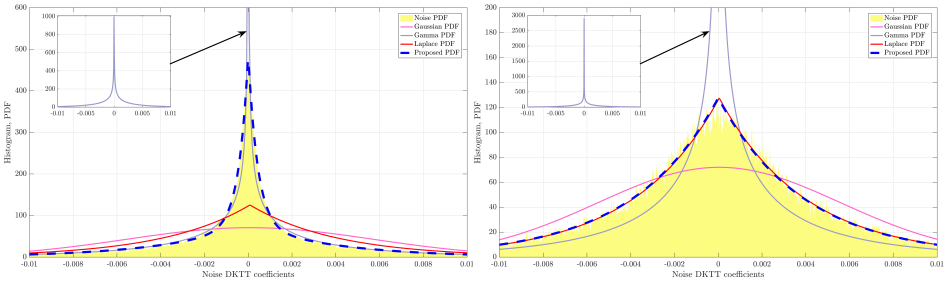
$$F_G\left(x_k, \sigma_{x_k}\right) = A_G \frac{e^{-\frac{|x_k|}{\sigma_{x_k}}}}{\sqrt{4\pi\sigma_{x_k} x_k}} \tag{30}$$

The definition of Laplacian density is

$$F_L\left(x_k, b_{x_k}\right) = \frac{1}{2b_{x_k}} e^{-\frac{|x_k|}{C_L\, b_{Lx_k}}} = \frac{1}{2b_{x_k}} e^{-\frac{|x_k|}{b_{x_k}}} \tag{31}$$

where the Laplacian factor is defined as $b_{x_k} = C_L\, b_{Lx_k}$, and the Laplacian variance is defined as $\sigma_L^2 = 2b_{x_k}^2$. The noise model ($F_{\text{Noise}}$) is assumed to be a combination of two Laplacian PDFs, as follows:

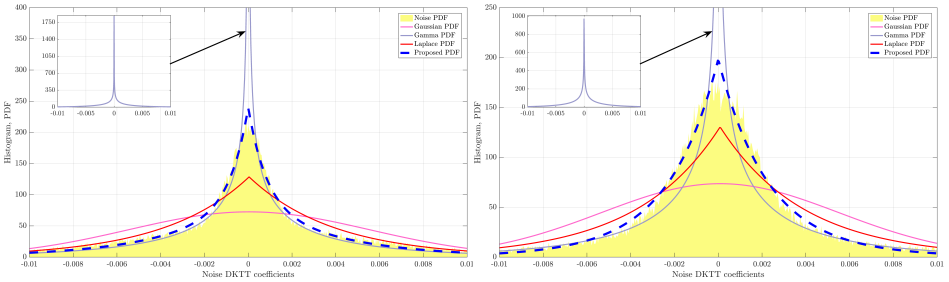
$$\begin{aligned} F_{Noise}\left(d_k, b_{d_{ki}}\right) &= \sum_{i=1}^{m} \frac{B_{di} e^{-\frac{|y_k - x_k|}{b_{d_{ki}}}}}{C_{di}\, b_{d_k}} \\ &= \sum_{i=1}^{m} \frac{B_{di} e^{-\frac{|y_k - x_k|}{b_{d_{ki}}}}}{b_{d_{ki}}} \end{aligned} \tag{32}$$




(a) Babble Noise



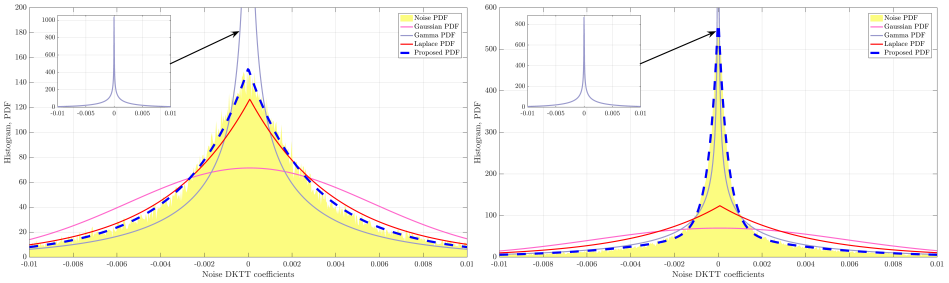

(b) Buccaneer Noise




(c) Destroyer Engine Noise



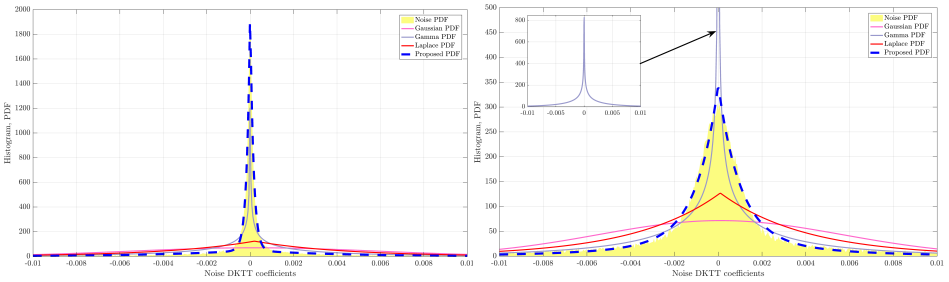

(d) Destroyer operations Noise




(e) F16 Noise



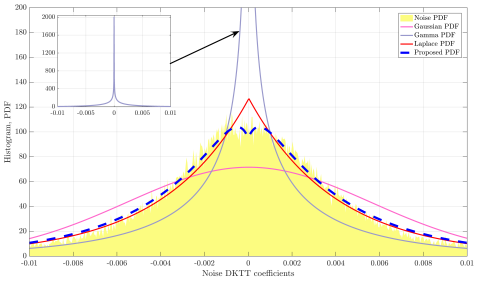

(f) Factory Noise




(g) Leopard Noise




(h) M109 Noise




(i) Pink Noise

FIGURE 4: The proposed PDF for different types of Noise.

Meanwhile, $(b_{d_{\text{ki}}} = C_{di}\ b_{d_k})$ represents the ith Laplacian factor, and the variance of the Laplacian noise PDF is $\sigma^2_{L_{\text{ki}}} = 2b^2_{d_{\text{ki}}}$. The mathematical formula for the NBSE estimator in a constructive interference event is

$$E\left[X_k|Y_k, E_+\right] = \frac{\int_{-\infty}^{\infty} x_k p\left(x_k, Y_k, E_+\right)}{p\left(Y_k, E_+\right)} dx_k \tag{33}$$

where $x_k$ and $y_k$ respectively represent the instances of random processes $X_k$ and $Y_k$. The same equation as (19) is obtained for $E_-$.The joint PDF of two independent random variables can be expressed by multiplying their marginal probability. Then, the joint PDF between $x_k$ and $y_k$ is [22], [62]

$$p\left(x_k, y_k, E_+\right) = \begin{cases} p_{XY}\left(x_k y_k\right) = p\left(x_k\right) p\left(y_k\right) & m_k Y_k > |X_k| \\ 0 & otherwise \end{cases} \tag{34}$$

where $m_k = sgn\ (X_k)$, $(F_{Speech})$ , and $(\ F_{Noise})$ are independent with a zero mean. When the long term of $E\left(X_k/Y_k, E_+\right)$ is considered after substituting $F_{Speech}$ and $F_{Noise}$, this term is divided into four parts, i.e., two for the numerator and two for the denominator. Then, the conditional expectation operator for constructive interference can be defined as

$$E\left(X_k/Y_k, E_+\right) = \frac{\int_0^{Y_k} x_k \left(A_G \frac{e^{-\frac{|x_k|}{\sigma_{x_k}}}}{\sqrt{4\pi\sigma_{x_k} x_k}} + A_L \frac{e^{\left(-\frac{|x_k|}{b_{x_k}}\right)}}{2b_{x_k}}\right)\left(\sum_{i=1}^{2} \frac{B_{d_i} e^{-\frac{|y_k - x_k|}{b_{d_{\text{ki}}}}}}{b_{d_{\text{ki}}}}\right) dx_k}{\int_0^{Y_k} \left(A_G \frac{e^{-\frac{|x_k|}{\sigma_{x_k}}}}{\sqrt{4\pi\sigma_{x_k} x_k}} + A_L \frac{e^{\left(-\frac{|x_k|}{b_{x_k}}\right)}}{2b_{x_k}}\right)\left(\sum_{i=1}^{2} \frac{B_{d_i} e^{-\frac{|y_k - x_k|}{b_{d_{\text{ki}}}}}}{b_{d_{\text{ki}}}}\right) dx_k} \tag{35}$$

The terms of the numerator $N_c$ are divided into $N_{c1}$ and $N_{c2}$. $N_{c1}$ is defined as follows:

$$N_{c1} = \int_0^{Y_k} x_k \left(A_G \frac{e^{-\frac{x_k}{\sigma_{x_k}}}}{\sqrt{4\pi\sigma_{x_k} x_k}}\right)\left(\sum_{i=1}^{2} \frac{B_{di} e^{-\frac{(y_k - x_k)}{b_{d_{ki}}}}}{b_{d_{\text{ki}}}}\right) dx_k \tag{36}$$

After solving the aforementioned integral using Theorem (3.381(1)) from [63] and simplifying it in terms of priori and posterior SNRs, the result is expressed in terms of an incomplete gamma function as follows:

$$N_{c1} = \sum_{i=1}^{2} \frac{B_{di} A_G}{4\sqrt{\pi}} \frac{e^{-\sqrt{\frac{\gamma_k}{C_{di}}}}}{\sqrt{\frac{\xi_k . c_G}{C_{di}}}} \left(\sqrt{\frac{C_{di}}{\xi_k . c_G}} - 1\right)^{-1.5} \Upsilon\left(1.5, \sqrt{\frac{\gamma_k}{C_{di}}}\left(\sqrt{\frac{C_{di}}{\xi_k . c_G}} - 1\right)\right) \tag{37}$$

The same mathematical solution is applied to $N_{C2}$, and the result is

$$\sum_{i=1}^{2} \frac{A_L\ B_{di}}{4} \left[\frac{\left(-\sqrt{\gamma_k \xi_k}\frac{\sqrt{C_L}}{C_{di}} + \sqrt{\xi_k}\sqrt{\frac{C_L}{C_{di}}} + \sqrt{\frac{\gamma_k}{C_{di}}}\right) e^{\sqrt{\frac{\gamma_k}{C_{di}}}} - \sqrt{\xi_k}\sqrt{\frac{C_L}{C_{di}}} e^{\sqrt{\frac{\gamma_k}{\xi_k}}\frac{1}{\sqrt{C_L}}}}{\left(1 - \sqrt{\xi_k}\sqrt{\frac{C_L}{C_{di}}}\right)^2}\right] \times e^{-\sqrt{\frac{\gamma_k}{C_{di}}}\left(1 + \frac{1}{\sqrt{\xi_k}}\sqrt{\frac{C_{di}}{C_L}}\right)} \tag{38}$$

The denominator $D_c$ is also separated into two terms. The first term $D_{c1}$ is expressed in terms of an incomplete gamma function, as follows:

$$\left(\sum_{i=1}^{2} \frac{B_{di} A_G}{4\sqrt{\pi}} \frac{e^{-\sqrt{\frac{\gamma_k}{C_{di}}}}}{\sqrt{\frac{\xi_k . c_G}{C_{di}}}} \left(\sqrt{\frac{C_{di}}{\xi_k . c_G}} - 1\right)^{-0.5} \times \Upsilon\left(0.5, \sqrt{\frac{\gamma_k}{C_{di}}}\left(\sqrt{\frac{C_{di}}{\xi_k . c_G}} - 1\right)\right)\right) \frac{1}{y_k} \tag{39}$$

The formula for the second part, $D_{c2}$, of the denominator will be

$$D_{c2}=\left(\sum_{i=1}^{2}\frac{A_L\,B_{di}}{4}\,\frac{\left(e^{\sqrt{\frac{\gamma_k}{\xi_k}}\frac{1}{\sqrt{C_L}}}-e^{\sqrt{\frac{\gamma_k}{C_{di}}}}\right)}{\left(1-\sqrt{\xi_k}\,\sqrt{\frac{C_L}{C_{di}}}\right)}e^{-\sqrt{\frac{\gamma_k}{C_{di}}}\left(1+\frac{1}{\sqrt{\xi_k}}\sqrt{\frac{C_{di}}{C_L}}\right)}\right)\frac{1}{y_k} \tag{40}$$

Finally, the general form of the speech estimator, $(\mathrm{NBSE})_c$, in a constructive event is

$$E\left(\frac{X_k}{Y_k},E_+\right)=(NBSE)_c=\left(\frac{N_{c1}+N_{c2}}{D_{c1}+D_{c2}}\right).y_k \tag{41}$$
$$=G_c^{NBSE}.y_k$$

From the other extreme, the analytical solution for NBSE in a destructive event is

$$E\left(\frac{X_k}{Y_k},E_-\right)=\frac{\int_{-\infty}^{0}x_k p_{\mathrm{XY}}\left(x_k,Y_k\right)dx_k+\int_{Y_k}^{\infty}x_k p_{\mathrm{XY}}\left(x_k,Y_k\right)dx_k}{\int_{-\infty}^{0}p_{\mathrm{XY}}\left(x_k,Y_k\right)dx_k+\int_{Y_k}^{\infty}p_{\mathrm{XY}}\left(x_k,Y_k\right)dx_k} \tag{42}$$

The destructive equations are clearly longer than the constructive equations; therefore, they are divided into eight parts, i.e., four for the numerator and four for the denominator. The first integral in (42) is termed as $N_{df}=N_{d1}+N_{d2}$, where $N_{d1}$ is

$$N_{d1}=\sum_{i=1}^{2}\frac{B_{di}A_G}{4\sqrt{\pi}}\frac{e^{-\sqrt{\frac{\gamma_k}{C_{di}}}}}{\sqrt{\frac{\xi_k.c_G}{C_{di}}}}\left(\sqrt{\frac{C_{di}}{\xi_k.c_G}}+1\right)^{-1.5}\Gamma(1.5) \tag{43}$$

The mathematical solution for the second term, $N_{d2}$, is calculated as follows:

$$N_{d2}=\sum_{i=1}^{2}-\frac{A_L\,B_{di}}{4}\,\frac{\left(\sqrt{\frac{\xi_k c_L}{C_{di}}}e^{-\sqrt{\frac{\gamma_k}{C_{di}}}}\right)}{\left(1+\sqrt{\frac{\xi_k.c_L}{C_{di}}}\right)^2} \tag{44}$$

Then, the second integral in the numerator is taken, $N_{\mathrm{ds}}=N_{d3}+N_{d4}$, where $N_{d3}$ is

$$N_{d3}=\sum_{i=1}^{2}\frac{B_{di}A_G}{4\sqrt{\pi}}\frac{e^{-\sqrt{\frac{\gamma_k}{C_{di}}}}}{\sqrt{\frac{\xi_k.c_G}{C_{di}}}}\left(\sqrt{\frac{C_{di}}{\xi_k.c_G}}+1\right)^{-1.5}\Gamma\left(1.5,\left(\sqrt{\frac{C_{di}}{\xi_k.c_G}}+1\right)\sqrt{\frac{\gamma_k}{C_{di}}}\right) \tag{45}$$

and $N_{d4}$ is

$$N_{d4}=\sum_{i=1}^{2}\frac{A_L B_{di}}{4}\,\frac{\left(\frac{\sqrt{\gamma_k\xi_k c_L}}{C_{di}}+\sqrt{\frac{\xi_k c_L}{C_{di}}}+\sqrt{\frac{\gamma_k}{C_{di}}}\right)e^{-\sqrt{\frac{\gamma_k}{\xi_k C_L}}}}{\left(1+\sqrt{\frac{\xi_k.c_L}{C_{di}}}\right)^2} \tag{46}$$

The denominator, $D_d=D_{df}+D_{ds}$, is also separated into two terms, namely, $D_{df}$ and $D_{ds}$. The mathematical calculation of the first term, $D_{df}=D_{d1}+D_{d2}$ , is as follows:

$$D_{d1}=\left(\sum_{i=1}^{2}\frac{B_{di}A_G}{4\sqrt{\pi}}\frac{e^{-\sqrt{\frac{\gamma_k}{C_{di}}}}\sqrt{\frac{\gamma_k}{C_{di}}}}{\sqrt{\frac{\xi_k.c_G}{C_{di}}}}\times\left(\sqrt{\frac{C_{di}}{\xi_k.c_G}}+1\right)^{0.5}\Gamma(0.5)\right)\frac{1}{y_k} \tag{47}$$

The second part $D_{d2}$ of the first part in $D_{df}$ has the following form:

$$D_{d2}=\left(\sum_{i=1}^{2}\frac{A_L\,B_{di}}{4}\,\frac{\left(e^{-\sqrt{\frac{\gamma_k}{C_{di}}}}\right)}{\left(1+\sqrt{\frac{\xi_k.c_L}{C_{di}}}\right)}\sqrt{\frac{\gamma_k}{C_{di}}}\right)\frac{1}{y_k} \tag{48}$$

The second term, $D_{ds}=D_{d3}+D_{d4}$, in the denominator. The equation for $D_{ds}$ is

$$D_{ds}=\int_{Y_k}^{\infty}x_k\left(F_{Speech}\right)\left(F_{Noise}\right)dx_k \tag{49}$$

The formula for $D_{d3}$ is

$$\left[\sum_{i=1}^{2}\frac{B_{di}A_G}{4\sqrt{\pi}}\frac{e^{\sqrt{\frac{\gamma_k}{C_{di}}}}\sqrt{\frac{\gamma_k}{C_{di}}}}{\sqrt{\frac{\xi_k.c_G}{C_{di}}}}\left(\sqrt{\frac{C_{di}}{\xi_k.c_G}}+1\right)^{-0.5}\Gamma\left(0.5,\left(\sqrt{\frac{C_{di}}{\xi_k.c_G}}+1\right)\sqrt{\frac{\gamma_k}{C_{di}}}\right)\right]\frac{1}{y_k} \tag{50}$$

The mathematical solution for the second term, $D_{d4}$, is

$$D_{d4} = \left( \sum_{i=1}^{2} \frac{A_L\, B_{di}}{4} \frac{\left( e^{-\sqrt{\frac{\gamma_k}{\xi_k C_L}}} \right)}{\left( 1 + \sqrt{\frac{\xi_k . c_L}{C_{di}}} \right)} \sqrt{\frac{\gamma_k}{C_{di}}} \right) \frac{1}{y_k} \quad (51)$$

Thus, the general form of the estimator in a destructive event $(NBSE)_d$ is

$$\begin{aligned} E\left[X_k | Y_k, E_-\right] &= (NBSE)_d \\ &= \left( \frac{N_{d1} + N_{d2} + N_{d3} + N_{d4}}{D_{d1} + D_{d2} + D_{d4} + D_{d4}} \right) . y_k \\ &= G_d^{NBSE} . y_k \end{aligned} \quad (52)$$

Then, Equation (13) of NBSE, which provides an optimal estimation of a clean signal, is

$$\hat{x}_k^{NBSE} = f_k \left(NBSE\right)_c + \left(1 - f_k\right) \left(NBSE\right)_d \quad (53)$$

### 2) The Proposed Linear Bilateral Super-Gaussian Estimator (LBSE)

To improve the performance of the speech enhancement process, the problem of residual noise, including MN, which is highly irritating to the human ears, must be addressed. Therefore, a post-processing filtering technique, i.e., LBSE, is proposed as a second stage estimator. Moreover, LBSE will deal with the over-attenuation problem in low SNR levels. The linear relation that combines $Y_k$ and $\hat{X}_k$ is expressed as

$$\hat{X}_k = G_k Y_k \quad (54)$$

where $G_k$ is the multiplicative LBSE gain. The expression for MSE has been defined previously. Then, the well-known expression for the linear MSE equation is written as follows [22], [29]:

$$e_k = E\left[ \left( X_k - G_k Y_k \right)^2 \right] \quad (55)$$

The general form of the multiplicative gain is derived as follows by differentiating and minimizing (41) with respect to the gain function and then equating it to zero:

$$G_k = \frac{E\left[{X_k}^2\right] + E\left[X_k D_k\right]}{E\left[{X_k}^2\right] + 2E\left[X_k D_k\right] + E\left[{D_k}^2\right]} \quad (56)$$

LBSE has two gains based on the relative term $E\left[X_K D_k\right]$. For a constructive event, the term is $E\left[|X_K|\,|D_K|\right] = E\left[|X_K|\right]\, \mathrm{E}\left[|D_K|\right]$; for a destructive event, the term is $E\left[|X_K|\,|D_K|\right] = -\,E\left[|X_K|\right]\, \mathrm{E}\left[|D_K|\right]$.

The cross term, $E\left[|X_K|\,|D_K|\right]$, plays an important role in determining the performance of a linear multiplicative gain filter. Equation (42) can be written in terms of $\xi_k$ and $p_E$ as follows:

$$G_k = \frac{\xi_k + p_E \sqrt{\xi_k}}{\xi_k + 2p_E \sqrt{\xi_k} + 1} \quad (57)$$

where $p_E$ is calculated as

$$p_E = \frac{\pm \left(A_G C_G + A_L C_L\right) \left(B_{N1} C_{d1} + B_{N2} C_{d2}\right)}{\sqrt{2}} \quad (58)$$

$p_E$ significantly affects noise reduction along with speech distortion. In LBSE, the same PDFs for speech and noise are assumed, and thus, the term $E\left[|X_K|\,|D_K|\right]$ must be calculated for these models. The expectation values of the speech signal $E\left[|X_K|\right]$ and noise signal $E\left[|D_K|\right]$ are [63]:

$$\begin{aligned} E\left[|X_K|\right] &= A_G . \frac{\sigma_{x_k}}{2} + A_L b_{x_k}, \\ E\left[|D_K|\right] &= B_{N1} . b_{d_{k1}} + B_{N2} b_{d_{k2}} \end{aligned} \quad (59)$$

When (45) is substituted into the cross term, the mathematical formulas for LBSE for constructive and destructive events are

$$G_c^{\text{LBSE}} = \frac{\xi_k + \frac{\left(A_G C_G + A_L C_L\right)\left(B_{N1} C_{d1} + B_{N2} C_{d2}\right)}{\sqrt{2}} \sqrt{\xi_k}}{\xi_k + 1 + \sqrt{2}\left(A_G C_G + A_L C_L\right) \times \left(B_{N1} C_{d1} + B_{N2} C_{d2}\right) \sqrt{\xi_k}} \quad (60)$$

$$G_d^{\text{LBSE}} = \frac{\xi_k - \frac{\left(A_G C_G + A_L C_L\right)\left(B_{N1} C_{d1} + B_{N2} C_{d2}\right)}{\sqrt{2}} \sqrt{\xi_k}}{\xi_k + 1 - \sqrt{2}\left(A_G C_G + A_L C_L\right) \times \left(B_{N1} C_{d1} + B_{N2} C_{d2}\right) \sqrt{\xi_k}} \quad (61)$$

Therefore, DCP plays a significant role in determining the optimum $p_E$ value. In FIGURE 5, the percentage value of MSE (equation (64)) is calculated to show the improvement of different $p_E$ values.

Since the term $E[X_k N_k]$ of LBSE is not zero in this work. Therefore, the formula of MSE-LBSE will be:

$$e_k^{LBSE} = (1 - G_k) E[X_k^2] - G_k E[X_k D_k] \quad (62)$$

In the meantime, $G_k$ has two events of noise interference; therefore, the probability of occurrence of each case in normal situation is assumed to be equally possible. Therefore, the general formula of $e_k^{LBSE}$ is

$$e_k^{LBSE} = \frac{E[X_k^2](\xi_k + 1)(1 - p_E^2)}{(\xi_{k+1})^2 - 2\xi_k (p_E^2)} \quad (63)$$

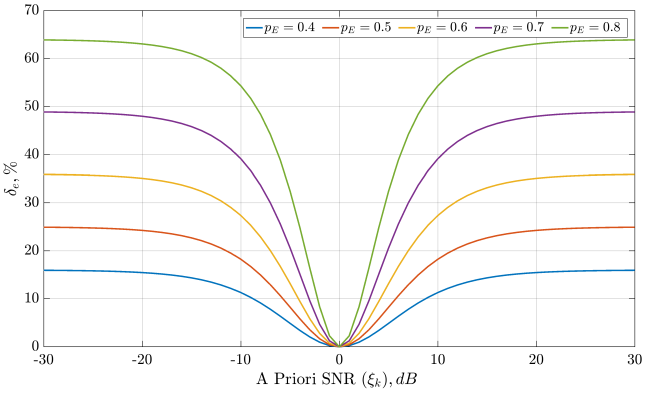



FIGURE 5: MSE Improvement of LBSE

The general formula for calculating the percentage of MSE improvement is

$$\delta_e = \frac{e_k^W - e_k^{LBSE}}{e_k^W} * 100\% \quad (64)$$

This equation is plotted as a function of $\xi_k$ to demonstrate the percentage of improvement between WF and the LBSE estimator for different $p_E$ values. Evidently, no improvement occurs when $\xi_k$ = 0. Meanwhile, the $\delta_e$ percentage of improvement begins to increase gradually as $p_E$ increases. The improvement in the proposed estimator reaches nearly 25% at $\xi_k$=±30 dB for $p_E = 0.5$ and 65% for $p_E = 0.8$. After estimating a clean speech signal, the inverse of DKTT is applied to convert the signal back to the time domain. The workflow of the proposed system is presented in FIGURE 6.

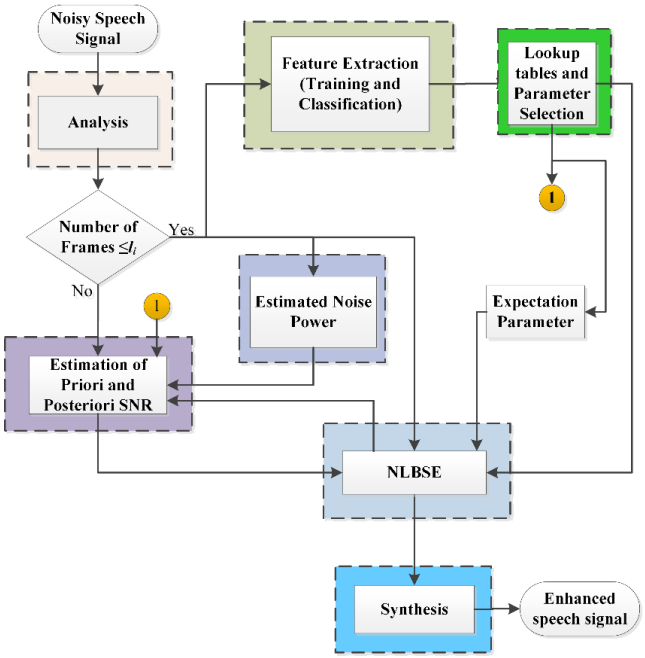



FIGURE 6: The Block Diagram of the proposed SEA.

## III. GAIN CHARACTERISTIC OF LBSE AND NBSE ESTIMATORS

In this section, the characteristics of the two proposed estimators are presented to illustrate their performance in filtering out unwanted components of a noise signal. For a constructive event, an attenuation estimator is required. For a destructive event, an amplification filter is required. The following sections present the characteristics of NBSE and LBSE. Each estimator has two gain formulas for each event.

### A. GAIN CHARACTERISTIC OF LBSE ESTIMATOR

In FIGURE 7a, various gain curves against $\xi_k$ for different values of $p_E$ are plotted in a destructive case. The gain formula for LBSE is a function of $\xi_k$. $p_E$ has different values because DCP has varying values.

Evidently, the gain value is equal to 0.5 for all values of $p_E$ when $\xi_k = 0dB$. By contrast, the highest curve in the region $\xi_k > 0dB$ is for $p_E = 0.8$, which amplifies the signal approximately after $\xi_k > 2dB$. For $\xi_k < 0dB$, the filter with $p_E = 0.5$ delivers less attenuation than the others. For both estimators, the curve gains become zero gain as $\xi_k$ approaches $\infty$ or $-\infty$. The figure clearly shows that gains do not always amplify noisy components as predicted for a destructive event. This counter-intuitive phenomenon can be elucidated if the occurrence of polarity reversal [22] is considered at a destructive event, particularly for regions where the gain has a negative value.

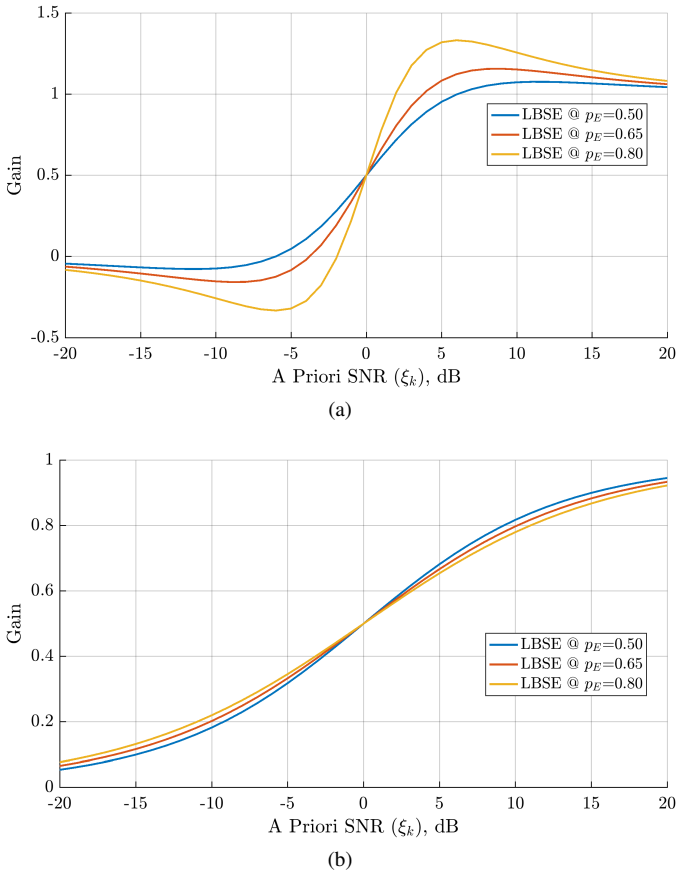



FIGURE 7: LBSE Gain Characteristic against a priori SNR for (a) destructive (b) constructive events.

In FIGURE 7b, the gain curves are plotted for a constructive case, $G_c^{\text{LBSE}}$. The plots are superimposed for better comparison. Evidently, when $\xi_k = 0dB$, the values of all the gains are equal to 0.5. In addition, for $\xi_k > 0dB$, the curve for $p_E = 0.8$ provides more attenuation to the signal, which is suitable for a constructive event, and vice versa. For both estimators, the curve gains tend toward zero gain as $\xi_k$ approaches $\infty$ or $-\infty$. Evidently, all gains are attenuation gains and less than the unity in all the regions of $\xi_k$. This property is appropriate for a constructive condition because the noise interference in such case always tends to increase noisy speech signals.

### B. GAIN CHARACTERISTIC OF NBSE ESTIMATOR

NBSE is a nonlinear estimator, the output of which is not linear with its input signal. NBSE is considerably harder to derive than LBSE. NBSE gain is a function of two parameters, namely, $\xi_k$ and $\gamma_k$. The 2D and 3D schemes of the

NBSE gain curves are plotted. Only white noise is presented due to space limitation. The NBSE gain function, $G_c^{\text{NBSE}}$, is plotted in FIGURE 8 as a function of $\xi_k$ for $\gamma_k = -10dB$ and $\gamma_k = 5dB$.

In general, the attenuation gain curves decrease gradually as $\xi_k$ decreases, which is good for maintaining signal distortion at an appropriate value. The 3D plot clearly shows that the gain of NBSE increases progressively within a little bit when $\gamma_k$ increases, thereby increasing the opportunity to improve the enhancement process. For a constructive case, the attenuation is low. Furthermore, it converges rapidly toward a higher gain as the value of $\gamma_k$ increases. NBSE provides an attenuation filtering gain in nearly all gain levels, which is significant for a constructive event.

FIGURE 9 shows the 3D and 2D plots of the parametric NBSE gain function, $G_d^{\text{NBSE}}$, when a destructive event is considered. In FIGURE 9a, the 3D plot of NBSE gain is shown based on the variation of two parameters, $\xi_k$ and $\gamma_k$. The 2D plot in FIGURE 9b shows the parametric gain curves as a function of $\xi_k$ for $\gamma_k = -10dB$ and $\gamma_k = 5dB$. For small values of $\gamma_k$, the gain becomes higher than the unity for the range of $\xi_k > -6\ dB$, as it should be for a destructive event.

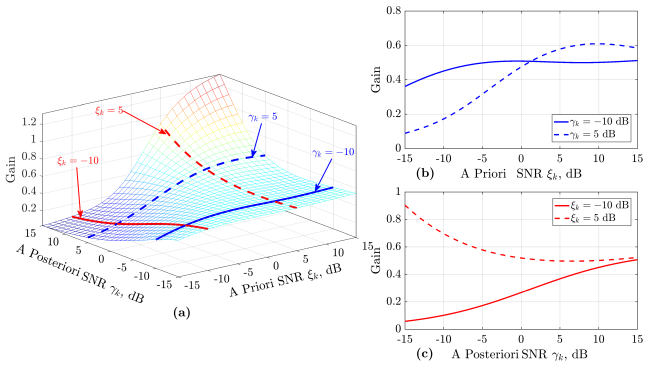



FIGURE 8: Gain curves of NBSE for white noise in constructive event.

However, when $\gamma_k$ increases, the NBSE filter tends to provide an attenuation gain that is appropriate for the case of polarity reversal, which may occur in this interference case. NBSE amplifies or attenuates each noise component in proportion to the estimated $\xi_k$ when $\gamma_k$ is constant. Interestingly, the gain levels are smaller than one given that $\xi_k$ is small, which causes attenuation in a degraded signal. However, gain value crosses the unity gain (0 dB) as $\xi_k$ increases to provide an amplification gain.

## IV. THE EVALUATION OF THE PROPOSED SEA.

An assessment of the proposed SEA is presented in the following sections.

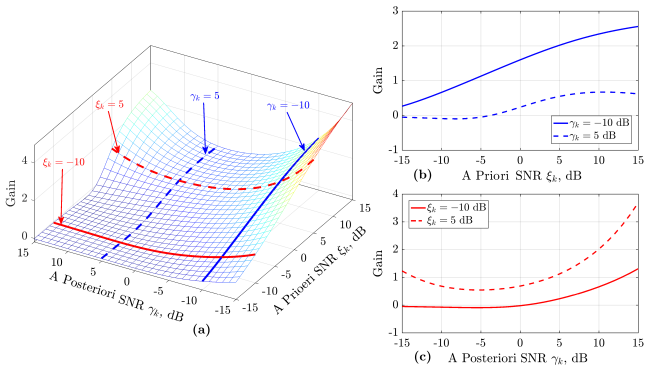



FIGURE 9: Gain curves of NBSE for white noise in destructive event.

### A. ACCURACY EVALUATION OF NOISE CLASSIFICATION METHOD

In the noise classification phase, 100 speech files are taken from the well-known TIMIT dataset [44] for the training phase, whereas 300 speech files are taken for the testing phase. The sampling rate is 16 KHz and 1-hamming window is used with 75% overlap. The speech files denoted by SA1 and SA2 are obtained for male and female speakers, respectively. 150 files for "SA1" and 150 files for "SA2". The speech signals are corrupted by eleven selected noise types. Among these, ten are selected from the NOISEX-92 dataset [64], in addition to the speech-shaped noise [44]. The types of noise include white, pink, F16, buccaneer, factory, babble, engine room noise, operation room noise, leopard, M109, and speech-shaped noise. Moreover, 5 levels of noise (-10, -5, 0, 5, 10 dB) are utilized for each noise type. The length of training and testing data for each SNR level is approximately 25 *ms*. Noise classification is carried out on the first six frames of the noisy speech. The features in moment domain are directly obtained from the noisy signal, where no other features are required to achieve a successful noise classification. Therefore, the complexity computational of classification process is low. After feature extraction process, the training stage is performed to gain the classifier model. This classifier model is used as a pre-stage before the process of SE to classify the 11 types of noises. The procedure of noise classification method can be summarized in the following steps:

Step 1: The input is a noisy signal (speech+ noise) from TIMIT [65] and Noisex-92 databases [64] of 400 $\times$ 11 $\times$ 5 speech files. These files consist of 400 speech files, 11 types of noise, and each speech file has five levels of SNR with frame size of 400 samples (25 *ms*).

Step 2: The initial six frames from each noisy file are taken to extract 50 features. These 50 features are contained 25 mean power features and 25 mean standard deviation features.

Step 3: The 22,000 speech files are divided into two sets which are training set (5500) and test set (16,500).

Meanwhile, the training set is treated by 5-fold cross validation.

Step 4: The training set is used to train the multi-class SVM. The parameters of the SVM are adjusted to make minimal the average error of 5-fold cross validation using grid search.

Step 5: The test dataset is constructed to analyze the performance of the classifier and then to calculate the confusion matrix. If acceptable, then output the classifier, otherwise return to step 4 to re-train the parameters of the SVM model.

The separation boundaries of different classes in SVM were determined by choosing the appropriate kernel function. As a reasonable choice, we adopted the polynomial kernel function with degree of two ($d = 2$ and $r = 1$) since this kernel has the lowest classification error against linear, radial basis function, and sigmoid kernels [66]. Its formula is as follows:

$$K(x_n, x_i) = (\gamma(x_n, x_i) + r)^d \tag{65}$$

The cross-validation and grid-search methods are used to tune the optimal kernel parameter ($\gamma$) and the penalty parameter ($C$). Where, cross-validation procedure can prevent the over fitting problem. In this work, the 5-fold cross validation is applied due to its simple and easy properties. The mechanism is to create a 5-fold partitions of the whole dataset. The dataset was partitioned into 5 disjoint, equal size subsets. The process is repeated 5 times to use 4 folds for training and a left fold for validation where, the test error was calculated, and finally average the validation error rates of 5 experiments. In each run, the best parameters of a classification algorithm for a class pair were explored through 5-fold cross validation with a grid search mechanism on the training set. The classifier with the least validation error was selected for each class pair.

The summary of the testing phase is provided in TABLE 4. The accuracy of the classification has been found to be 99.43%. For example, the percentage accuracy for Buccaneer noise (3rd class) is 99.87 %, and the percentage accuracy for factory noise to babble noise is 1.2. A low percentage accuracy is obtained for babble noise (2nd class), i.e., 97.60 %. The confusion matrix shows that the proposed noise classification method attains high accuracy in different noise environments.

The average accuracy of the proposed noise classification method for all the eleven types of noise is 99.43%. For example, the percentage accuracy for Buccaneer noise (3rd class) is 99.87, and the percentage accuracy for factory noise to babble noise is 1.2. A low percentage accuracy is obtained for babble noise (2nd class), i.e., 97.60. The confusion matrix shows that the proposed noise classification method attains high accuracy in different noise environments.

### B. PERFORMANCE EVALUATION OF NLBSE USING QUALITY AND INTELLIGIBILITY MEASURES

This section provides a performance assessment of the proposed SEA compared with several existing methods to establish its capability in suppressing noise perfectly. A comparative evaluation is used to assess speech intelligibility and quality. However, listening tests is a gold standard in terms of speech quality valuation; these tests are expensive and time-consuming, which limit their application [67]. Accordingly, powerful objective measures are adopted in the present study. The number of speech files used in this experimental test is 64, with different speakers (32 males and 32 females), which are randomly selected from the TIMIT database [65] to make the work complementary with mean opinion scores for hearing quality. The decision-directed approach [12] is implemented to compute the estimated $\xi_k$ with variable $\propto$ based on noise type as follows:

$$\hat{\xi}_l(k) = \propto \frac{\hat{X}_{l-1}(k)}{\hat{\lambda}_{D,l-1}(k)} + (1- \propto) max(\hat{\gamma}_l(k) - 1, 0) \tag{66}$$

The tests are performed on the eleven types of noise [64] with SNRs of -10, -5, 0, 5, 10 dB SNR. Then, five quality measures are used: PESQ [68], composite measures (SIG, BAG, and OVL) [69], [70], and FWSNR [67]. Two intelligibility measures are used, namely, CSII [71] and STOI [72]. A comprehensive assessment is performed on four selected classes of methods: (1) traditional estimators: WF [3] and the nonlinear MMSE estimator [12]; (2) low-distortion methods: dual-gain Wiener DGW [29], Laplacian–Gaussian mixture-based dual-gain Wiener filter (LGMDGW) [73], and dual MMSE estimator (DMMSE) [22]; (3) two-stage SEA using OP: two-stage-based DKT estimator [6] (TSDKTE) and two-stage-based DTT estimator [6] (TSDTTE); and (4) a recent method called the optimally modified log-spectral amplitude based on noise classification (COMLSA) [44].

Each speech signal is divided into frames with a length of 18 ms. The standard hamming window with 75% overlap is used for the framing process. The optimal value of parameter $p$ in DKTT transform is set to 0.2. For combination, the enhanced speech signal in each frame is synthesized via the overlap–add method [74]. White noise is selected as an example, as shown in FIGURE 10, to calculate quality and intelligibility measures. FIGURE 12 shows that NLBSE provides higher measurement values for all noisy conditions, except for SNR = 10 dB in FWSNR and SNR = 5 dB and 10 dB in CSII, where the NLBSE value is comparable with those of the other algorithms. In general, NLBSE provides the best results in low SNR levels for PESQ, SIG, BAK, and OVL. NLBSE is verified to have the highest value compared with the other selected methods.

PESQ is known for its high correlation with OVL measures, which in turn, exhibit a significant correlation with subjective speech quality [44]. Meanwhile, the speech-shaped noise in FIGURE 11 shows that NLBSE is better than all the other algorithms. The experimental results for the

TABLE 4: The confusion matrix of the noise classification method

| | | Predicted Classes | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Noise Type | White | Babble | F16 | Speech Shaped | Pink | Buccaneer | Destroy Engine | Destroy Operation | Leopard | M109 | Factory |
| Actual Classes | White | **100** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Babble | 0.00 | **97.60** | 0.07 | 0.07 | 0.00 | 0.00 | 0.00 | 0.13 | 0.00 | 0.33 | 1.80 |
| | F16 | 0.00 | 0.00 | **99.93** | 0.00 | 0.00 | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Speech Shaped | 0.00 | 0.00 | 0.00 | **100** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Pink | 0.00 | 0.00 | 0.00 | 0.00 | **100** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Buccaneer | 0.00 | 0.00 | 0.00 | 0.13 | 0.00 | **99.87** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | Destroy Engine | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **100** | 0.00 | 0.00 | 0.00 | 0.00 |
| | Destroy Operation | 0.00 | 0.20 | 0.07 | 0.00 | 0.07 | 0.00 | 0.00 | **99.33** | 0.00 | 0.27 | 0.07 |
| | Leopard | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **99.73** | 0.27 | 0.00 |
| | M109 | 0.00 | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.47 | 0.00 | **99.40** | 0.07 |
| | Factory | 0.00 | 1.20 | 0.27 | 0.00 | 0.13 | 0.00 | 0.00 | 0.07 | 0.00 | 0.40 | **97.93** |

other types of noise indicate that NLBSE provides the highest values in nearly all noise situations.

The amount of residual noise in the enhanced speech noise cannot be quantified easily by using only objective measures. However, spectrogram representations of an enhanced speech can be applied to provide additional details on the time–frequency distribution. FIGURE 12 shows the spectrogram plot of a speech sentence obtained from the TIMIT dataset that was corrupted by white noise with 0 dB SNR. The spectrogram of a noisy signal is shown in FIGURE 12b. The spectrograms of NLBSE and the other methods are displayed in this figure. The sentence used is, "She had your dark suit in greasy wash water all year" Clean and noisy spectrograms are also provided to perform comparison evaluation and confirm the optimal process of the proposed SEA. NBSE is also presented to prove the capability of LBSE. Evidently, a clean signal is regenerated using NLBSE without noticeable signal distortion and with minimum residual noise, where no noise surrounds the original signal in the spectrogram. Moreover, the spectrograms present how the opportunity to enhance a noisy signal is increased by utilizing the second post-processing filtering shown in FIGURE 12c. The imminent analysis of other algorithms will start with DGW and DMMSE. FIGURES 12f and 12g clearly show that residual noise, including MN, surrounds the original signal. COMLSA in FIGURE 12h shows that less residual noise appears as isolated peaks in the frequency domain. By contrast, the TSDTTE estimator efficiently removes residual noise. However, speech distortion is clearly shown in FIGURE 12e. Evidently, spectrogram view reinforces the capability of NLBSE to remove noise with less speech distortion and residual noise, including MN.

## V. CONCLUSION AND FUTURE WORK

This paper addresses significant problems of SEA in estimating a clean speech signal under different environments of background noise. The proposed SEA adopts a noise classification method, which is used to search for accurate speech and noise models. A new super-Gaussian composite is assumed and used first in modeling. Two stages of estimators are derived based on the models, namely, NBSE and LBSE, which are distinct from other estimators in terms of their analytical solution. These two estimators are based on a low-distortion approach and MMSE sense; they are then combined in cascade to realize NLBSE. NLBSE is proposed to minimize distortion under different conditions of the underlying speech signal during the enhancement process without compromising the noise reduction process. It is adopted

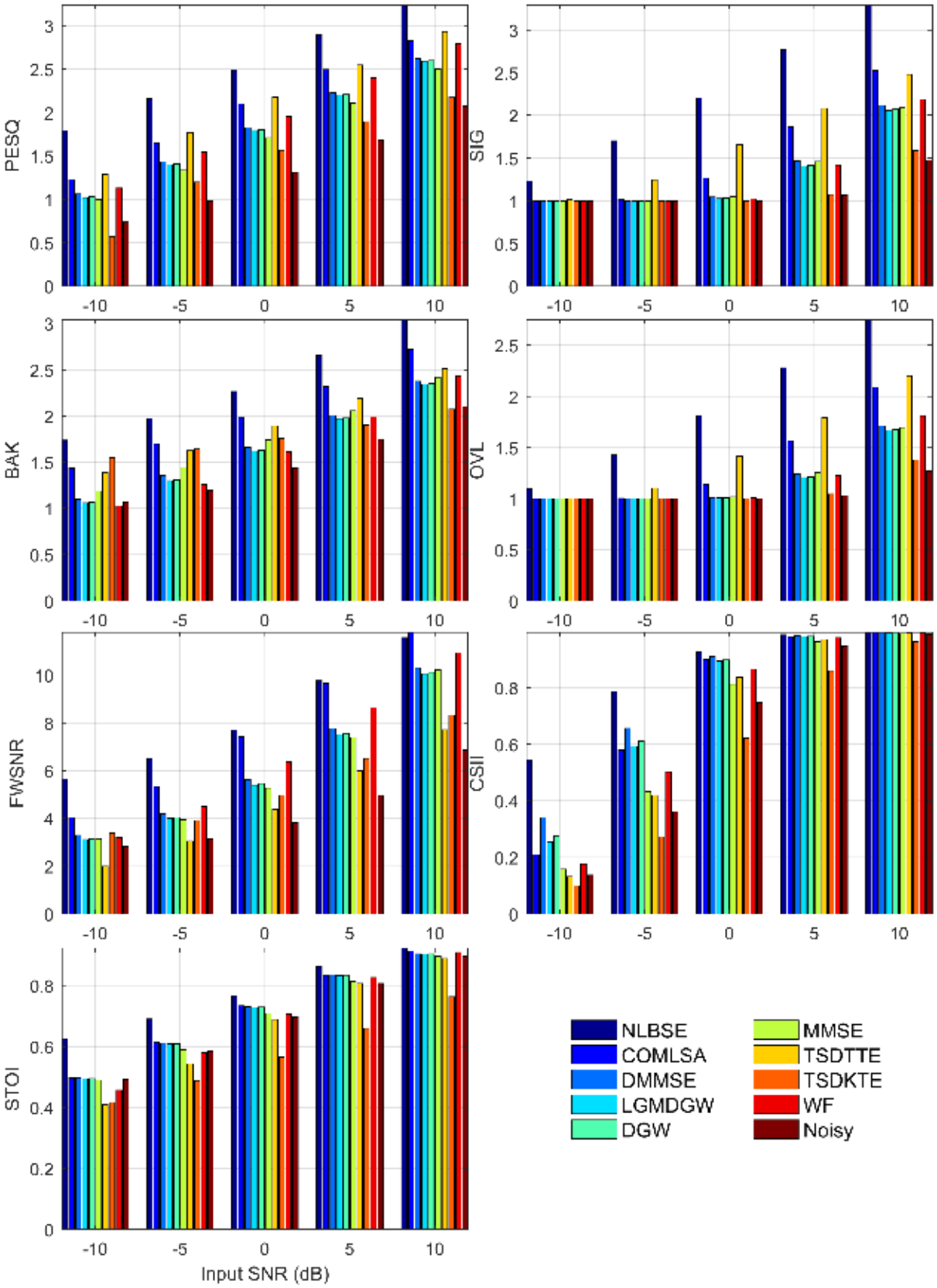



FIGURE 10: The Comparison test of white noise condition for seven measurements.

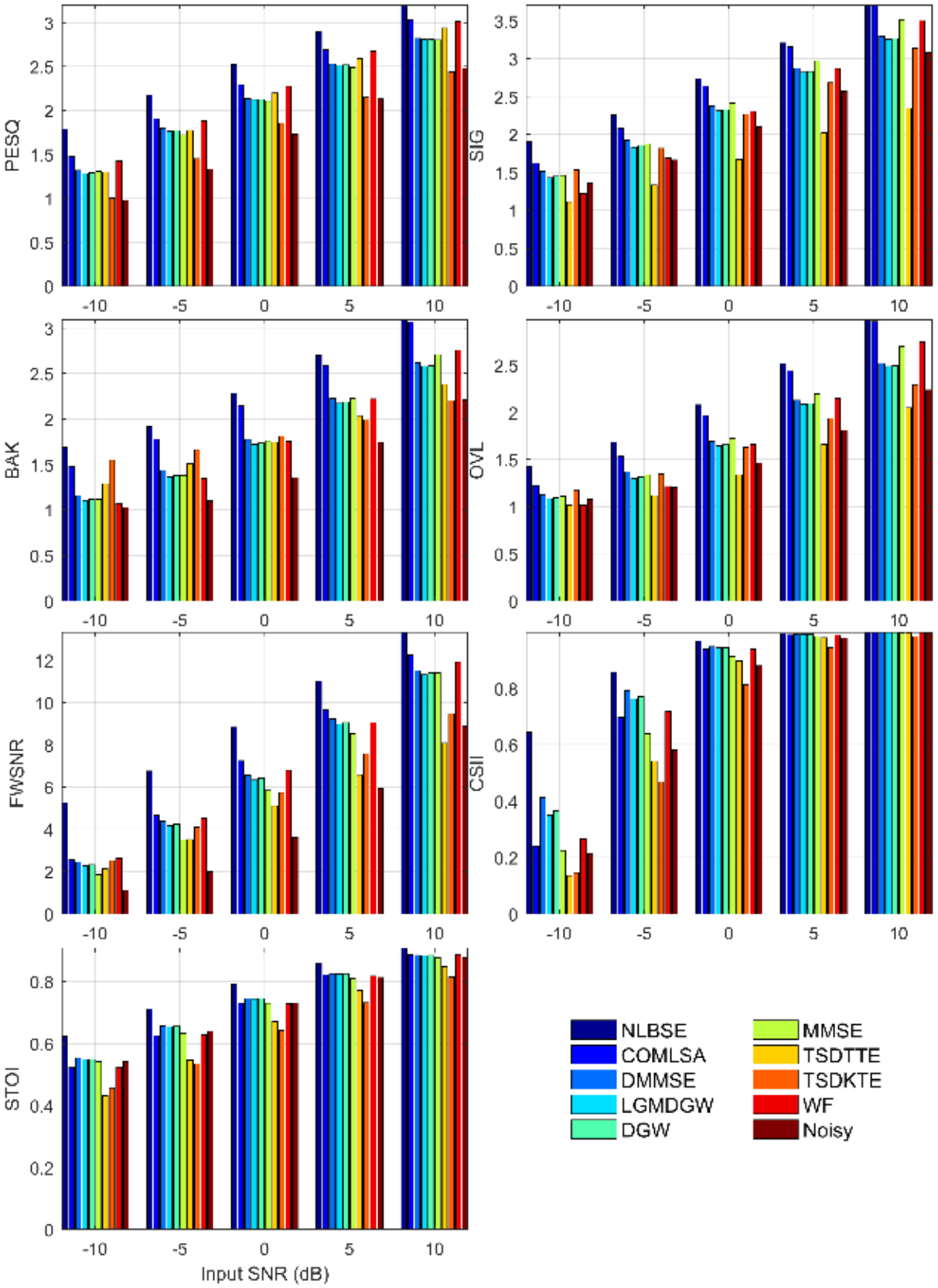
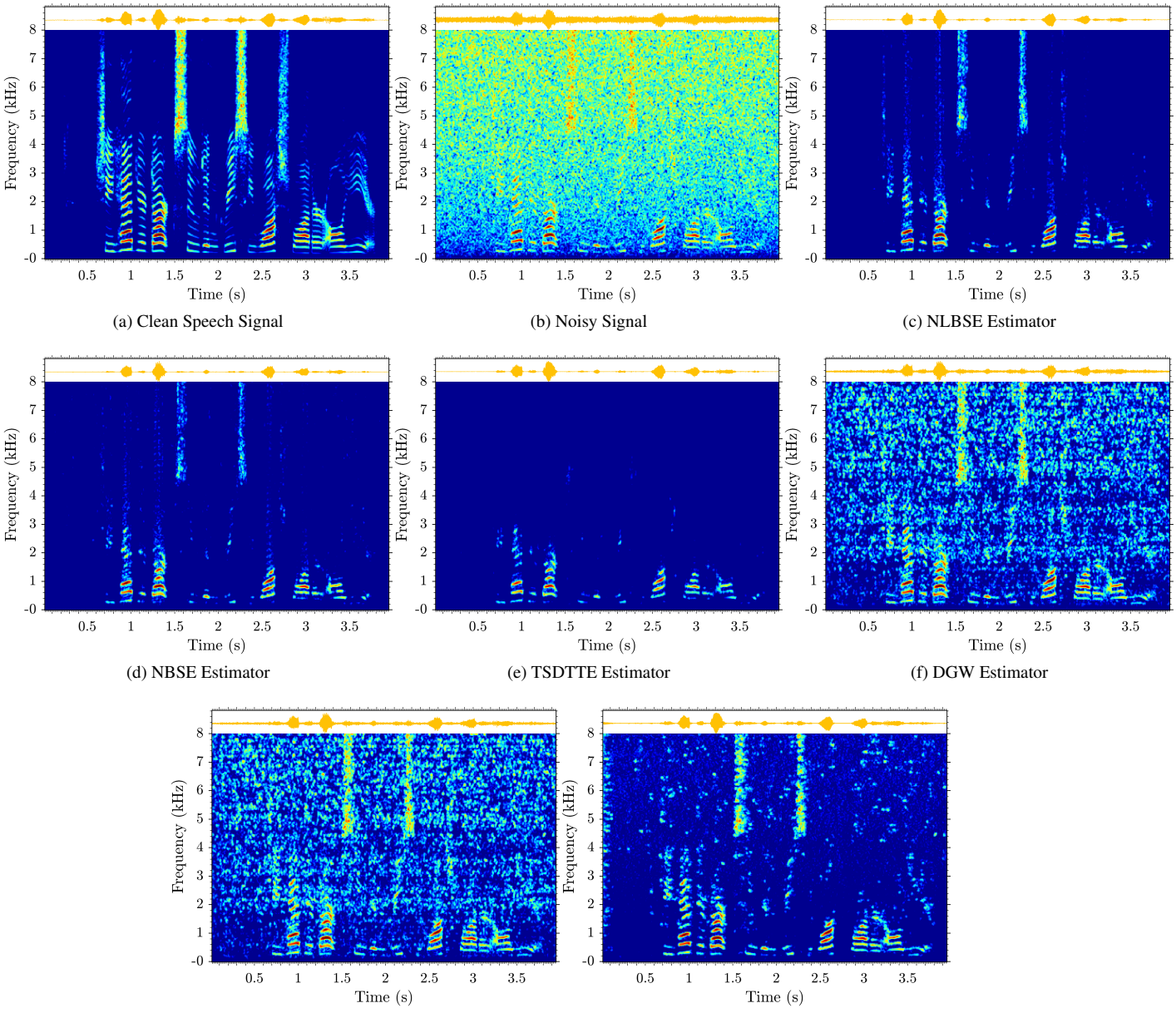



FIGURE 11: The Comparison test of Speech-shaped noise condition for seven measurements.




(a) Clean Speech Signal (b) Noisy Signal (c) NLBSE Estimator

(d) NBSE Estimator (e) TSDTTE Estimator (f) DGW Estimator

(g) DMMSE Estimator (h) COMLSA Estimator

FIGURE 12: Result of the enhancement process for the male utterance "She had your dark suit in greasy wash water all year" taken from the TIMIT database corrupted by white noise with 0 dB SNR. The spectrogram plots of (a) clean speech, and (b) noisy signals; and enhanced signals using (c) NLBSE, (d) NBSE, (e) TSDTTE, (f) DGW, (g) DMMSE, and (h) COMLSA.

by considering the interference between clean and noise signals and the type of noise. Only a few algorithms deal with these approaches. The proposed estimators address the polarity reversal issue that occurs when noise components are stronger than signal components. High-performance noise suppression is achieved from the NBSE output, with more enhancement for speech perceptual aspects besides to reduced MN effect in LBSE.

The analytical solutions of MMSE for linear and nonlinear estimators are derived. The outcomes of the proposed estimators demonstrate their effectiveness and capability to reduce unwanted noise in terms of different speech quality and intelligibility measures. The simulation results of different noisy conditions clearly show that the proposed work reduces corrupting noise in a degraded signal in a superior manner compared with various existing methods. In the future, the proposed work will be applied to calculate an optimum value for the polarity estimator factor in practical cases of bilateral gain. Furthermore, other types of super-Gaussian prior and noise will be examined.

## REFERENCES


[1] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," Proceedings of the IEEE, vol. 67, no. 12, pp. 1586–1604, 1979.

[2] X. Yousheng and H. Jianwen, "Speech enhancement based on combination of wiener filter and subspace filter," in Audio, Language and Image Processing (ICALIP), 2014 International Conference on. IEEE, Conference Proceedings, pp. 459–463.

[3] P. C. Loizou, Speech enhancement: theory and practice. CRC press, 2013.

[4] H. R. Abutalebi and M. Rashidinejad, "Speech enhancement based on $\beta$-order mmse estimation of short time spectral amplitude and laplacian speech modeling," Speech Communication, vol. 67, pp. 92–101, 2015.

[5] B. Liu, J. Tao, Z. Wen, and F. Mo, "Speech enhancement based on analysis-synthesis framework with improved parameter domain enhancement," Journal of Signal Processing Systems, vol. 82, no. 2, pp. 141–150, 2016. [Online]. Available: http://dx.doi.org/10.1007/s11265-015-1025-1

[6] W. A. Jassim, R. Paramesran, and M. S. Zilany, "Enhancing noisy speech signals using orthogonal moments," IET Signal Processing, vol. 8, no. 8, pp. 891–905, 2014.

[7] M. Zoulikha and M. Djendi, "A new regularized forward blind source separation algorithm for automatic speech quality enhancement," Applied Acoustics, vol. 112, pp. 192–200, 2016.

[8] S. Chehrehsa and T. J. Moir, "Speech enhancement using maximum a-posteriori and gaussian mixture models for speech and noise periodogram estimation," Computer Speech and Language, vol. 36, pp. 58–71, 2016.

[9] N. Mohammadiha, P. Smaragdis, and A. Leijon, "Supervised and unsupervised speech enhancement using nonnegative matrix factorization," IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 10, pp. 2140–2151, 2013.

[10] D. Wang and J. Chen, "Supervised speech separation based on deep learning: An overview," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 10, pp. 1702–1726, 2018.

[11] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 27, no. 2, pp. 113–120, 1979.

[12] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 32, no. 6, pp. 1109–1121, 1984.

[13] B. Chen and P. C. Loizou, "A laplacian-based mmse estimator for speech enhancement," Speech communication, vol. 49, no. 2, pp. 134–143, 2007.

[14] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," Speech and Audio Processing, IEEE Transactions on, vol. 3, no. 4, pp. 251–266, 1995.

[15] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," Speech and Audio Processing, IEEE Transactions on, vol. 11, no. 4, pp. 334–341, 2003.

[16] J. Lim, A. Oppenheim, and L. Braida, "Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 26, no. 4, pp. 354–358, 1978.

[17] J. Tu and Y. Xia, "Effective kalman filtering algorithm for distributed multichannel speech enhancement," Neurocomputing, vol. 275, pp. 144–154, 2018.

[18] Y. Xia and J. Wang, "Low-dimensional recurrent neural network-based kalman filter for speech enhancement," Neural Networks, vol. 67, pp. 131–139, 2015.

[19] J. Tu and Y. Xia, "Fast distributed multichannel speech enhancement using novel frequency domain estimators of magnitude-squared spectrum," Speech Communication, vol. 72, pp. 96–108, 2015.

[20] S. Abdulhussain, A. Ramli, M. Saripan, B. Mahmmod, S. Al-Haddad, and W. Jassim, "Methods and Challenges in Shot Boundary Detection: A Review," Entropy, vol. 20, no. 4, p. 214, mar 2018. [Online]. Available: http://www.mdpi.com/275270 http://www.mdpi.com/1099-4300/20/4/214

[21] Y. Soon, S. N. Koh, and C. K. Yeo, "Noisy speech enhancement using discrete cosine transform," Speech communication, vol. 24, no. 3, pp. 249–257, 1998.

[22] T. Hasan and M. K. Hasan, "Mmse estimator for speech enhancement considering the constructive and destructive interference of noise," IET signal processing, vol. 4, no. 1, pp. 1–11, 2010.

[23] S. I. Yann, "Transform based speech enhancement techniques," PhD Thesis, Nanyang Technological University, Singapore, 2003.

[24] H. Ding, Y. Soon, and C. K. Yeo, "A dct-based speech enhancement system with pitch synchronous analysis," IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 8, pp. 2614–2623, 2011.

[25] D. L. Donoho, "De-noising by soft-thresholding," IEEE transactions on information theory, vol. 41, no. 3, pp. 613–627, 1995.

[26] B. M. Mahmmod, A. R. Ramli, S. H. Abdulhussain, S. A. R. Al-Haddad, and W. Jassim, "Signal Compression and Enhancement Using a New Orthogonal-Polynomial-Based Discrete Transform," IET Signal Processing, vol. 12, no. 1, pp. 129–142, aug 2018.

[27] T. Gerkmann, M. Krawczyk-Becker, and J. Le Roux, "Phase processing for single-channel speech enhancement: History and recent advances," IEEE Signal Processing Magazine, vol. 32, no. 2, pp. 55–66, 2015.

[28] S. Samui, I. Chakrabarti, and S. K. Ghosh, "Improved single channel phase-aware speech enhancement technique for low signal-to-noise ratio signal," IET Signal Processing, vol. 10, no. 6, pp. 641–650, 2016.

[29] I. Soon and S. Koh, "Low distortion speech enhancement," IEE Proceedings-Vision, Image and Signal Processing, vol. 147, no. 3, pp. 247–253, 2000.

[30] S. Gazor and W. Zhang, "Speech enhancement employing laplacian-gaussian mixture," IEEE Transactions on Speech and Audio Processing, vol. 13, no. 5, pp. 896–904, 2005.

[31] R. Martin, "Speech enhancement using mmse short time spectral estimation with gamma distributed speech priors," in Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on, vol. 1. IEEE, Conference Proceedings, pp. I–253–I–256.

[32] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 33, no. 2, pp. 443–445, 1985.

[33] R. Martin, I. Wittke, and P. Jax, "Optimized estimation of spectral parameters for the coding of noisy speech," in Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on, vol. 3. IEEE, Conference Proceedings, pp. 1479–1482.

[34] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," Signal processing, vol. 81, no. 11, pp. 2403–2418, 2001.

[35] B. Kirubagari, S. Palanivel, and N. Subathra, "Speech enhancement using minimum mean square error filter and spectral subtraction filter," in Information Communication and Embedded Systems (ICICES), 2014 International Conference on. IEEE, Conference Proceedings, pp. 1–7.

[36] E. Ambikairajah, G. Tattersall, and A. Davis, "Wavelet transform-based speech enhancement," in Fifth International Conference on Spoken Language Processing, 1998, Conference Proceedings.

[37] E. Jayakumar and P. Sathidevi, "Speech enhancement based on noise type and wavelet thresholding the multitaper spectrum," in Advances in Machine Learning and Signal Processing. Springer, 2016, pp. 187–200.

[38] P. J. Wolfe and S. J. Godsill, "Efficient alternatives to the Ephraim and Malah suppression rule for audio signal enhancement," EURASIP Journal on Advances in Signal Processing, vol. 2003, no. 10, p. 910167, 2003.

[39] X. Zou and X. Zhang, "Speech enhancement using an mmse short time dct coefficients estimator with supergaussian speech modeling," Journal of Electronics (China), vol. 24, no. 3, pp. 332–337, 2007.

[40] B. M. Mahmmod, A. R. Ramli, S. H. Abdulhussain, S. Al-Haddad, and W. A. Jassim, "Low-distortion mmse speech enhancement estimator based on laplacian prior," IEEE Access, vol. 5, no. 1, pp. 9866–9881, 2017.

[41] S. Gazor and Z. Wei, "Speech probability distribution," IEEE Signal Processing Letters, vol. 10, no. 7, pp. 204–207, 2003.

[42] R. Martin, "Statistical methods for the enhancement of noisy speech," Speech Enhancement, pp. 43–65, 2005.

[43] R. Martin and C. Breithaupt, "Speech enhancement in the dft domain using laplacian speech priors," in Proc. IWAENC, vol. 3, Conference Proceedings, pp. 87–90.

[44] W. Yuan and B. Xia, "A speech enhancement approach based on noise classification," Applied Acoustics, vol. 96, pp. 11–19, 2015.

[45] M. Kolbk, Z.-H. Tan, and J. Jensen, "Speech intelligibility potential of general and specialized deep neural network based speech enhancement systems," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 1, pp. 153–167, 2017.

[46] C. C. E. de Abreu, M. A. Q. Duarte, and F. V. Alvarado, "An immunological approach based on the negative selection algorithm for real noise classification in speech signals," AEU-International Journal of Electronics and Communications, vol. 72, pp. 125–133, 2017.

[47] R. Li, Y. Liu, Y. Shi, L. Dong, and W. Cui, "ILMSAF based speech enhancement with DNN and noise classification," Speech Communication, vol. 85, pp. 53–70, 2016.

[48] B. Xia and C. Bao, "Wiener filtering based speech enhancement with weighted denoising auto-encoder and noise classification," Speech Communication, vol. 60, pp. 13–29, 2014.

[49] S. H. Abdulhussain, A. R. Ramli, B. M. Mahmmod, M. I. Saripan, S. A. R. Al-Haddad, and W. A. Jassim, "Shot boundary detection based on orthogonal polynomial," Multimedia Tools and Applications, pp. 1–22, feb 2019. [Online]. Available: http://link.springer.com/10.1007/s11042-019-7364-3

[50] S. H. Abdulhussain, "Temporal Video Segmentation Using Squared Form of Krawtchouk-Tchebichef Moments," PhD, Universiti Putra Malaysia, 2018.

[51] S. H. Abdulhussain, A. R. Ramli, S. A. R. Al-Haddad, B. M. Mahmmod, and W. A. Jassim, "On computational aspects of tchebichef polynomials for higher polynomial order," IEEE Access, vol. 5, pp. 2470–2478, 2017.

[52] G. Boros, V. H. Moll, and J. Foncannon, "Irresistible integrals: symbolics, analysis and experiments in the evaluation of integrals," The Mathematical Intelligencer, vol. 28, no. 3, pp. 65–68, 2006.

[53] S. H. Abdulhussain, A. R. Ramli, B. M. Mahmmod, M. I. Saripan, S. Al-Haddad, and W. A. Jassim, "A New Hybrid form of Krawtchouk and Tchebichef Polynomials: Design and Application," Journal of Mathematical Imaging and Vision, pp. 1–16, nov 2018. [Online]. Available: http://link.springer.com/10.1007/s10851-018-0863-4

[54] S. H. Abdulhussain, A. R. Ramli, S. A. R. Al-Haddad, B. M. Mahmmod, and W. A. Jassim, "Fast recursive computation of krawtchouk polynomials," Journal of Mathematical Imaging and Vision, vol. 60, no. 3, pp. 285–303, 2018.

[55] S. H. Abdulhussain, A. R. Ramli, B. M. Mahmmod, S. Al-Haddad, and W. A. Jassim, "Image edge detection operators based on orthogonal polynomials," International Journal of Image and Data Fusion, vol. 8, no. 3, pp. 293–308, 2017.

[56] H. S. Radeaf, B. M. Mahmmod, S. H. Abdulhussain, and D. Al-Jumaeily, "A steganography based on orthogonal moments," in Proceedings of the International Conference on Information and Communication Technology - ICICT '19, ser. ICICT '19. New York, New York, USA: ACM Press, 2019, pp. 147–153.

[57] S. H. Abdulhussain, A. R. Ramli, A. J. Hussain, B. M. Mahmmod, and W. A. Jassim, "Orthogonal polynomial embedded image kernel," in Proceedings of the International Conference on Information and Communication Technology - ICICT '19, ser. ICICT '19. New York, New York, USA: ACM Press, 2019, pp. 215–221. [Online]. Available: http://doi.acm.org/10.1145/3321289.3321310 http://dl.acm.org/citation.cfm?doid=3321289.3321310

[58] C. Lim, S.-R. Lee, and J.-H. Chang, "Efficient implementation of an SVM-based speech/music classifier by enhancing temporal locality in support vector references," IEEE Transactions on Consumer Electronics, vol. 58, no. 3, pp. 898–904, 2012.

[59] J. Milgram, M. Cheriet, and R. Sabourin, "âĂIJone against oneâĂİ or âĂIJone against allâĂİ: Which one is better for handwriting recognition with svms?" in Tenth international workshop on frontiers in handwriting recognition. Suvisoft, 2006.

[60] T. H. Al Banna, "A hybrid speech enhancement method using optimal dual gain filters and emd based post processing," Thesis, 2008.

[61] P. Vary and R. Martin, Digital speech transmission: Enhancement, coding and error concealment. John Wiley and Sons, 2006.

[62] A. Papoulis and S. U. Pillai, Probability, random variables, and stochastic processes. Tata McGraw-Hill Education, 2002.

[63] A. Jeffrey and D. Zwillinger, Table of integrals, series, and products. Academic Press, 2007.

[64] A. Varga and H. J. Steeneken, "Assessment for automatic speech recognition: Ii. noisex-92: A database and an experiment to study the effect of additive noise on speech recognition systems," Speech communication, vol. 12, no. 3, pp. 247–251, 1993.

[65] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, "Timit acoustic-phonetic continuous speech corpus," Linguistic data consortium, Philadelphia, 1993.

[66] M. Achirul Nanda, K. Boro Seminar, D. Nandika, and A. Maddu, "A comparison study of kernel functions in the support vector machine and its application for termite detection," Information, vol. 9, no. 1, p. 5, 2018.

[67] P. C. Loizou, Speech quality assessment. Springer, 2011, pp. 623–654.

[68] I. REC, "P. 862," Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, 2001.

[69] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 1, pp. 229–238, 2008.

[70] Y. Hu and p. C. Loizou, "Evaluation of objective measures for speech enhancement," in Ninth International Conference on Spoken Language Processing, 2006, Conference Proceedings.

[71] J. M. Kates and K. H. Arehart, "Coherence and the speech intelligibility index," The Journal of the Acoustical Society of America, vol. 117, no. 4, pp. 2224–2237, 2005.

[72] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 7, pp. 2125–2136, 2011.

[73] J. Wei, S. Ou, S. Shen, and Y. Gao, "Laplacian-gaussian mixture based dual-gain wiener filter for speech enhancement," in 2016 IEEE International Conference on Signal and Image Processing (ICSIP), Conference Proceedings, pp. 543–547.

[74] R. E. Crochiere, "A weighted overlap-add method of short-time fourier analysis/synthesis," Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 28, no. 1, pp. 99–102, 1980.

[75] A. K. E. Mizel, "Orthogonal functions solving linear functional differential equationsusing chebyshev polynomial," Baghdad Science Journal, vol. 5, no. 1, pp. 143–148, 2008.

BASHEERA M. MAHMMOD was born in Baghdad, Iraq, in 1975. She received the B.Sc. in Electrical Engi- neering in 1998 from Baghdad University. Then, she proceeded with her Master Degree in Electronic and Communication Engineering / Computer from Baghdad Universi ty in 2012. In 2018, she received the PhD degree in Computer and Embedded System Engineering from Universiti Putra Malaysia. Since 2007 until now, she is staff member at department of Computer Engineering, Faculty of Engineering, University of Baghdad. Her research interests include speech enhancement, signal processing, computer vision, RFID, and cryptography.

ABD RAHMAN RAMLI received a Bachelor of Science in Electronics, Universiti Kebangsaan Malaysia in 1982. Then he proceeded with his Master Degree in Information Technology System at the University of Strath- clyde, United Kingdom in 1985. In 1990, he pursued his doctoral studies in University of Bradford, United Kingdom. He was appointed as the Head of Computer and Communication System Engineering in August 1996 until July 1998. Abd Rahman had also served as Head of Intelligent Systems and Robotics Labora- tory, Institute of Advanced Technology, Universiti Putra Malaysia where he leads a cutting edge research laboratory in Real-Time and Embedded Systems, Intelligent Systems and Perceptual Robotics. His research interests are in the area of image processing and electronic imaging, multimedia systems engineering, embedded system and intelligent systems.

THAR BAKER Thar Baker received the Ph.D. degree in autonomic cloud applications from Liverpool John Moores University, U.K, in 2010, where he is currently a Senior Lecturer in distributed systems engineering and the Head of the Applied Computing Research Group, Faculty of Engineering and Technology. He became a Senior Fellow of the Higher Education Academy, in 2018. He has published numerous refereed research papers in multidisciplinary research areas including: big data, algorithm design, green and sustainable computing, and SDN. He has been actively involved as a member of editorial board and review committee for a number peer reviewed international journals (e.g., Future Generation Computer Systems journal), and is on programme committee for a number of international conferences. He is an Expert Evaluator of EU H2020, ICTFund, and British Council.

FERAS AL-OBEIDAT Feras Al-Obeidat is an Assistant Professor at the College of Technological Innovation at Zayed University. He received both his Masters and Ph.D. in Computer Science from the University of New Brunswick, Canada. His primary field of research is Data Mining and Machine Learning. Directly following his Ph.D., Dr. Al-Obeidat contributed to industrial, university and government teaching and research with premier organizations including IBM Canada.

SADIQ H. ABDULHUSSAIN was born in Baghdad, Iraq, in 1976. He received the B.Sc. and M.Sc. degrees in Electrical Engineering from Baghdad University, in 1998 and 2001, respectively. In 2018, he received the PhD degree in Computer and Embedded System Engineering from Universiti Putra Malaysia. Since 2005 until now, he is staff member at department of Computer Engineering, Faculty of Engineering, University of Baghdad. His research interests include computer vision, signal processing, speech and image processing.

WISSAM A. JASSIM was born in Baghdad, Iraq, in 1976. He received the B.Sc. and M.Sc. degrees in Electrical Engineering from Baghdad University, in 1999 and 2002, respectively, and the Ph.D. degree in Electrical Engineering from the University of Malaya, Kuala Lumpur, Malaysia in 2012. From 2013 to 2015, he was a Visiting Research Fellow with the Department of Biomedical Engineering, University of Malaya, Kuala Lumpur, Malaysia. From 2015 to 2016, he was a Post-Doctoral Fellow with the Department of Electrical Engineering, University of Malaya, Kuala Lumpur, Malaysia. He is currently a Research Fellow with the ADAPT Center, School of Engineering, Trinity College Dublin, the University of Dublin, Dublin 2, Ireland. His current research interests include machine learning, speech, and image processing.