

Received January 23, 2020, accepted February 7, 2020, date of publication February 19, 2020, date of current version March 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2975069

# Single-Channel Speech Enhancement Based on Adaptive Low-Rank Matrix Decomposition

CHAO LI, TING JIANG<sup>ID</sup>, SHENG WU<sup>ID</sup>, AND JIANXIAO XIE

Key Laboratory of Universal Wireless Communication, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Sheng Wu (thuraya@bupt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61671075, and in part by the Major Program of National Natural Science Foundation of China under Grant 61631003.

**ABSTRACT** The low-rank matrix decomposition (LMD) algorithm based on the maximum correntropy criterion (MCC) has recently shown its superiority to other algorithms in classification (e.g., face recognition), and we develop it into single-channel speech enhancement for the low-rank structure of speech signals in the time domain. However, a new issue has arisen: some residual noise exists in the enhanced speech due to its sensitivity to the exact rank value. To address this issue, we propose a novel adaptive LMD (ALMD) algorithm in which the energy threshold technique is adopted to adaptively update the effective rank value of each frame of the speech matrix. Our proposed ALMD algorithm can achieve an acceptable performance for low signal-to-noise ratio (SNR) levels without approximating the speech phase with the noisy phase. We compare ALMD algorithm with common conventional algorithms in Gaussian white noise and non-Gaussian noise conditions. The simulation results demonstrate that ALMD algorithm can achieve its superiority in terms of the segmental SNR (segSNR), perceptual evaluation of speech Quality (PESQ), and short-time objective intelligibility measure (STOI), when compared with tested baseline algorithms.

**INDEX TERMS** Maximum correntropy criterion, single-channel speech enhancement, adaptive low-rank matrix decomposition, energy threshold technique.

## I. INTRODUCTION

With the development of speech technology, many speech applications have emerged, such as smart phones, hearing aid, and human-computer interaction. These applications play indispensable roles in human life [1], [2]. However, they are often severely damaged by ubiquitous background noise. Therefore, the design of an appropriate speech enhancement algorithm is of great importance. To address this challenging issue, many solutions have emerged during the last fifty years. According to the number of microphones, these approaches could be broadly divided into two patterns: multi-channel speech enhancement and single-channel speech enhancement [3], [4]. Multi-channel speech enhancement requires a series of microphones to get the desired signal information [5], [6]. In this case, the spatial diversity is mainly exploited to separate target speech signal from mixed signals. Beamforming is one of the most commonly technology to perform spatial filtering. It forms a beam of the desired direction speech

signal through the array geometry. While zero trapping to the non-target directional interference source [2], [7]. However, the limitation for beamforming is whose reliance on accurate target source orientation information and acoustic transfer functions (ATFs).

Single-channel speech enhancement has attracted plenty of attention since it requires only one microphone to provide a more computationally appealing solution [8], [9]. Spectral subtraction (SS) [10] is one of the most widely used methods due to its simplicity, which was originally presented by Boll [11]. The enhanced speech is mainly obtained by subtracting the estimated noise energy from the noisy power, where the estimated noise power is detected by voice activity detection (VAD). The major drawback is that it suffers from an uncomfortable signal called “music noise”. Wiener filtering (WF) [12] is a typical filtering algorithm. The main idea of WF is that a filter function based on the minimizing estimate error criterion is used to get the target speech [13]. It is sufficiently robust to Gaussian additive noise but non-Gaussian noise. An alternative popular work is that several based minimum mean square error (MMSE) [14] schemes

The associate editor coordinating the review of this manuscript and approving it for publication was Stavros Ntalampiras<sup>ID</sup>.

have been proposed, such as MMSE short-time spectral amplitude estimator (MMSE-STSA) [15], MMSE-spectrum power estimator based on zero cross-terms (MMSE-SPZC) [16], and MMSE log spectral amplitude estimator (MMSE-log) [17]. They perform well to improve speech intelligibility and quality, especially in some noisy conditions. However, these MMSE-based methods are computationally expensive and their performance is sensitive to the prior knowledge of noise distribution. Another more promising work is the subspace approach, which decomposes the noisy speech into the noise subspace and the speech subspace via employing the Karhunen-Loeve transform (KLT) or singular value decomposition (SVD) [18], [19]. It is well applied with the assumption of subspace of speech and noise being orthogonal, which often fail in actual scenarios [20].

In recent years, a new matrix decomposition theory called robust principal component analysis (RPCA) [21]–[25] was developed for speech enhancement. In RPCA, the observation matrix is effectively separated as a low-rank component and a sparse component, representing speech and additive noise, respectively. Because of the temporal variability of the low-rank component, RPCA can perform well with nonstationary noise [24]. However, traditional RPCA has two drawbacks. The first drawback is that reasonable performance requires prior knowledge of the signal source. The other is that RPCA exhibits high computational complexity because it uses a singular value decomposition (SVD) process for large matrices. Up to the present, various methods have been attempted to alleviate these problems [23], [26]–[29]. Riemannian robust principal component pursuit (R2PCP) [30] is one well-known improved RPCA algorithm, which imposes the tailored Riemannian optimization function to avoid full SVD. However, it assumes that the rank value is fixed and the iteration starting point is random. Constrained low-rank and sparse matrix decomposition (CLSMD) [23] is a notable new method for RPCA-based speech enhancement. The core idea of CLSMD solution is to incorporate rank and sparsity constraints into the decomposition of noisy. It successfully deals with residual noise with a fast convergence speed. However, approximation of the speech phase is required, because it works in the time-frequency domain. Moreover, [31] has shown that an inaccurate phase decreased the quality of speech communication. Is there a new approach without the high computational cost of SVD and the phase processing is not required?

For the purpose of comfortable voice communication, we develop a novel single-channel speech enhancement method based on an adaptive low-rank matrix decomposition (ALMD) method inspired by the low-rank matrix decomposition (LMD) algorithm [32]. In [32], the maximum correntropy criterion (MCC) [33] is adopted to eliminate outlier effects, whilst the combination of half-quadratic (HQ) [34] optimization and greedy bilateral (GreB) [35] paradigm are used to accelerate the computational speed. Therefore, the approach not only avoids SVD but also displays state-of-the-art performance on face recognition. In addition, in the time domain, the speech signal can be considered a low-rank

component because of its short-time stability, whereas the background noise can be considered as the unknown corruption. Consequently, it was possible for us to develop the LMD algorithm for speech enhancement. However, a new issue has arisen that it fails to obtain the desired results for denoising due to its sensitivity to the exact rank value. Here, the energy threshold method is exploited to adaptively update the effective rank value of clean speech.

The excellent properties of our work could be concluded as follows:

- To our knowledge, the MCC-based LMD algorithm is the first time developed into single-channel speech enhancement for the low-rank structure of speech signal.
- The proposed method mainly exploits the low-rank structure of speech signals for matrix decomposition in the time domain. As a result, the accurate VAD process and noise estimation are needless herein.
- Since the presented algorithm only performs denoising in the time domain, approximating the speech phase by the noisy phase is not required. Therefore, it will be more conducive to real-time denoising issue.
- Owing to the adaptive estimate of the effective rank value, great performance can be obtained for the task of obtaining the low rank speech components. Experimental results reveal that the ALMD approach can perform significantly better than several common baseline methods for various different types of noise (white, pink, babble, F16, and hfchannel), especially in low signal-to-noise (SNR) situations.

The organization of this paper is listed as below. In section II, the relevant theoretical basis of RPCA theory is briefly introduced. Then, speech enhancement based on ALMD approach is described in section III. In section IV, we compare the adopted method with existing ones and analyse the numerical results. Finally, Section V concludes our work.

## II. RPCA THEORY

Principal component analysis (PCA) [36], [37] is one of well-known statistical methods for signal processing, can accurately recover the desired matrix with low dimension from high-dimensional observations. However, the performance of the PCA technique will be seriously degraded until large noise or outliers occurs. To address this robustness issue, a modified version called RPCA [38] was originally presented by J.Wright. The core idea is that a given noisy signal  $\mathbf{Y} \in \mathbb{R}^{N \times K}$  could be effectively divided into two parts: sparse  $\mathbf{S} \in \mathbb{R}^{N \times K}$  and low-rank  $\mathbf{L} \in \mathbb{R}^{N \times K}$  matrices (i.e.,  $\mathbf{Y} = \mathbf{L} + \mathbf{S}$ ). Owing to the speech signal has a sparse structure and uncorrelated white noise has a low-rank structure in time-frequency domain. Thus, the RPCA approach could be developed into speech enhancement. In enhancement process of RPCA approach,  $\mathbf{S}$  and  $\mathbf{L}$  are utilized to represent the speech and noise components, respectively. The convex optimization problem of RPCA can be described as,

$$\min \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1, \quad \text{s.t. } \mathbf{Y} = \mathbf{L} + \mathbf{S}, \quad (1)$$

where the symbol  $\lambda > 0$  denotes a weighting parameter, which is a significant factor to trade off the sparsity of  $\mathbf{S}$  with the low-rankness of  $\mathbf{L}$ . Also,  $\|\cdot\|_*$  and  $\|\cdot\|_1$  is defined as the nuclear norm and the  $l_1$ -norm of a matrix, respectively. Here, the nuclear norm and the  $l_1$ -norm is served as a surrogate for the rank of  $\mathbf{L}$  matrix and the sparsity of  $\mathbf{S}$ , respectively. To address the optimization problem in (1), different kinds of efficient algorithms have been presented, such as augmented Lagrangian method (ALM) [39], alternating direction method of multipliers (ADMM) [21] and accelerated proximal gradient (APG) [40]. However, since the delicate balance between the sparsity and low rankness may be not satisfied, the RPCA-based approach is fail to get the desired result.

One of the notable work is called CLSMD [23], which can effectively solve the RPCA problems by incorporating the constraints of low-rankness and sparsity. The noisy matrix could be described as  $\mathbf{Y} = \mathbf{L} + \mathbf{S} + \mathbf{O}$ , and the symbol  $\mathbf{O}$  represents a residual noise matrix. Before applying the CLSMD method to recover  $\mathbf{S}$  and  $\mathbf{L}$ , the spectral magnitude matrix of noisy is smoothed as  $\hat{\mathbf{Y}}$ . It can be written as

$$|\hat{\mathbf{Y}}(t, k)| = \frac{1}{3} (|\mathbf{Y}(t - 1, k)| + |\mathbf{Y}(t, k)| + |\mathbf{Y}(t + 1, k)|), \quad (2)$$

where the  $t$  and  $k$  is denoted as time index and frequency bin of  $\mathbf{Y}$ , respectively. Then, the CLSMD is performed to deal with the two sub-problems: recovering  $\mathbf{S}$  and  $\mathbf{L}$  from  $\hat{\mathbf{Y}}$ . It can be defined as

$$\begin{aligned} L_{(n)} &= \arg \min \|\hat{\mathbf{Y}} - \mathbf{L} - \mathbf{S}_{(n-1)}\|_F^2, \\ S_{(n)} &= \arg \min \|\hat{\mathbf{Y}} - \mathbf{L}_{(n)} - \mathbf{S}\|_F^2, \\ \text{s.t. } \text{rank}(\mathbf{L}) &\leq r = \min(t, k), \text{Card}(\mathbf{S}) \leq c \& \mathbf{S} \geq 0, \end{aligned} \quad (3)$$

where  $n$  represents the number of iteration.  $\|\cdot\|_F$  denotes the Frobenius norm (i.e.,  $\|\mathbf{E}\|_F = \sqrt{e_{ij}^2}$ ). Besides,  $\text{rank}(\cdot)$  is the rank function of estimated noise matrix, which is usually set to 1 or 2, due to the strong correlation of the column vectors of the background noise.  $\text{Card}(\cdot)$  is equivalent to  $\|\cdot\|_1$ , which is used to calculate the cardinality value of speech matrix. The drawback of CLSMD approach is that it adopts the noisy phase to approximate the clean speech phase and uses a hard threshold operator to achieve the acquisition of  $\mathbf{S}$  and  $\mathbf{L}$ .

### III. SPEECH ENHANCEMENT BASED ON ALMD

In this section, we will introduce ALMD-based speech enhancement, which can alleviate the problems of CLSMD and RPCA. More specifically, we first give the problem formulation and then describe the ALMD method.

#### A. PROBLEM FORMULATION

The task of speech enhancement will be considered in this section. Specifically, the process of acquiring an enhanced speech is performed in the time domain not the time frequency domain, which can address the shortcomings of the

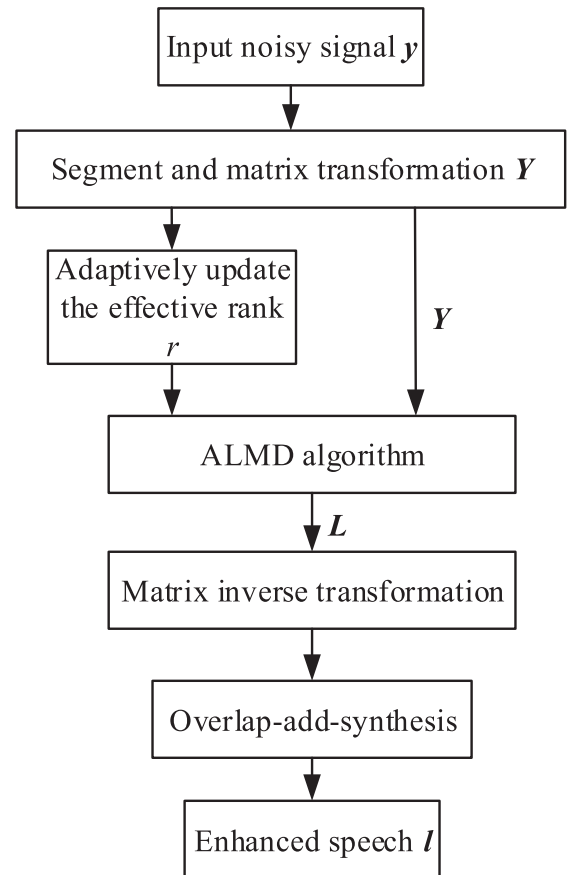


FIGURE 1. The ALMD-based speech enhancement framework.

approximate speech phase in the CLSMD method. However, owing to the stationary in the short time features of the speech signal, the noisy needs to divide the signal into  $N$  frames  $(\mathbf{y}_n, n = 1, \dots, N)$ . Let the symbol  $\mathbf{y}(t) \in \mathbb{R}^{K \times 1}$  denotes a frame of the noisy signal vector, which is taken to be the sum of a clean speech vector  $\mathbf{I}(t) \in \mathbb{R}^{K \times 1}$  and a noise vector  $\mathbf{s}(t) \in \mathbb{R}^{K \times 1}$ , where the symbol  $K$  is the length of each frame signal. The additive model can be given as

$$\mathbf{y}(t) = \mathbf{I}(t) + \mathbf{s}(t). \quad (4)$$

To show the low-rank characteristics of  $K$ -dimensional signals more clearly, a matrix transformation is performed on each frame of the signal. It can easily be obtained as

$$\mathbf{Y} = \mathbf{L} + \mathbf{S}, \quad (5)$$

where  $\mathbf{Y}$ ,  $\mathbf{L}$  and  $\mathbf{S}$  are  $M \times P$  matrices,  $M = 2K/3 + 1$ , and  $P = K/3$ . In the case of enhancement, the matrix  $\mathbf{Y}$  is known, while the  $\mathbf{L}$  and  $\mathbf{S}$  matrices are unknown. The ALMD algorithm is utilized to obtain speech matrix  $\mathbf{L}$ . After obtaining the estimated  $\mathbf{L}$ , an overlap-add-synthesis is utilized to recover the enhanced speech signal. This process is depicted in Fig. 1.

#### B. OPTIMIZATION ALGORITHM FOR ALMD

The core idea of ALMD is motivated by the MCC-based LMD method. However, since the speech energy of each

frame is unique, it is obvious that it is not feasible to use the fixed rank value. Thus, we will describe the adaptive rank value estimate at the end.

### 1) MCC AND HQ OPTIMIZATION

Correntropy [41] is a local criterion of similarity, which is often used to process the non-Gaussian noise with large outliers. Meanwhile, it also as a nonlinear similarity evaluation of two random variables  $\bar{\mathbf{X}}$  and  $\mathbf{G}$ . Moreover, since the correntropy holds the symmetrical characteristics. It can be defined as

$$V_{\sigma}(\bar{\mathbf{X}}, \mathbf{G}) = V_{\sigma}(\mathbf{G}, \bar{\mathbf{X}}) = \mathbb{E}[\kappa_{\sigma}(\bar{\mathbf{X}} - \mathbf{G})], \quad (6)$$

where the function  $\kappa_{\sigma}(\cdot)$  and  $\mathbb{E}[\cdot]$  denotes the kernel function and the expectation operator, respectively. Besides,  $\sigma$  is the kernel size that can be chosen by the maximum likelihood of density estimate. We will only consider the data  $\bar{x}$  that obeys a Gaussian distribution, so the  $\kappa_{\sigma}(\cdot)$  is the Gaussian kernel, i.e.,

$$\kappa_{\sigma}(\bar{x}) = \exp\left(-\frac{\bar{x}^2}{\sigma^2}\right). \quad (7)$$

In practice, since only the initial data  $\{(\bar{x}_i, g_i)\}_{i=1}^n$  are known but the joint probability density function (PDF), a simple estimator of correntropy could be used to instead of the expectation. It can be expressed as

$$\hat{V}_{n,\sigma}(\bar{\mathbf{X}}, \mathbf{G}) = \frac{1}{n} \sum_{i=1}^n \kappa_{\sigma}(\bar{x}_i - g_i). \quad (8)$$

Owing to the locality of correntropy, the similarity of the two variables (i.e.,  $\bar{x} = g$ ) is primarily dictated by  $\kappa_{\sigma}(\bar{x} - g)$  [41]. Therefore, a new cost function is called MCC that used to evaluate adaptive systems training. Giving as

$$\max_{\theta} \frac{1}{n} \sum_{i=1}^n \kappa_{\sigma}(e_i), \quad (9)$$

where the symbol  $\theta$  represents a set of adjustable parameters, and  $e_i$  is errors of system. Furthermore, the M-estimator [33] is also a maximum likelihood method, defined as  $\min_{\theta} \frac{1}{n} \sum_{i=1}^n (1 - \kappa_{\sigma}(e_i))$ . In that sense, M-estimator is equal to MCC. To address the problem of M-estimator, we make use of the HQ [34], [42] optimization.

HQ technique is a commonly used optimization algorithm for convex or non-convex minimization, which is well used for signal and image reconstruction [42]. The HQ function is defined as

$$Q_1(a, b) = (a - b)^2, \quad (10)$$

$$Q_2(a, b) = \frac{1}{2}a^2b, \quad (11)$$

where  $a$  and  $b$  is the adjustable parameter and an auxiliary parameter of the adaptive systems, respectively.  $b$  can be obtain by

$$b = \delta_a(a) = a \left(1 - \exp\left(-\frac{a^2}{\sigma^2}\right)\right), \quad (12)$$

where  $\delta(\cdot)$  is a minimization function in Welsch M-estimator. Owing to the additive form in it, (10) is well utilized to recover corrupted data [34]. Here, we only consider this case. Let us define a loss function of  $a$ , i.e.,

$$\psi(a) = \min_b Q_1(a, b) + \varphi(b), \quad (13)$$

where  $\varphi(\cdot)$  is served as the dual function of  $\psi(\cdot)$  in (13). MCC and HQ optimization could well deal with the problem of LMD, which is introduced in the following section.

### 2) MCC-BASED LMD APPROACH

To address another issue for the hard threshold operator in the CLSMD method. The MCC theory is adopted to provide an additional penalty for the updating of the LMD algorithm. Moreover, the method can also perform well in non-Gaussian noise conditions because of the property of correntropy [41]. Therefore, the redefined noise  $\mathbf{W} = \mathbf{S} + \mathbf{O}$  is modeled by the MCC theory, i.e.,

$$\max_{\mathbf{L}} \sum_{m=1}^M \sum_{p=1}^P \kappa_{\sigma}(\mathbf{W}_{mp}), \text{ s.t. } \mathbf{Y} = \mathbf{L} + \mathbf{W}, \text{ rank}(\mathbf{L}) \leq r, \quad (14)$$

where the  $\kappa_{\sigma}(\cdot)$  is a Gaussian kernel defined in (7). Based on the HQ optimization approach, an auxiliary variable  $\mathbf{G}$  is introduced, and  $\varphi(\mathbf{G})$  serves as the dual function of  $\psi(\mathbf{W})$  (i.e.,  $\psi(\mathbf{W}) = \psi(\mathbf{Y} - \mathbf{L}) = \min_{\mathbf{L}, \mathbf{G}} (\mathbf{Y} - \mathbf{L} - \mathbf{G})^2 + \varphi(\mathbf{G})$ ). Thus,  $\mathbf{G}$  can be easily obtained by (12) and (7), and it is given as

$$\mathbf{G} = \mathbf{W} - \mathbf{W} \odot \kappa_{\sigma}(\mathbf{W}), \quad (15)$$

where the mathematical symbol  $\odot$  indicates the entry-wise product. Therefore, Sub-problems of alternative updating for the MCC-based LDM approach in the CLSMD (3) solution can be rewritten as

$$\begin{aligned} \mathbf{L}_{(n)} &= \arg \min \|\mathbf{Y} - \mathbf{L} - \mathbf{G}_{(n-1)}\|_F^2 + \varphi(\mathbf{G}), \\ \mathbf{G}_{(n)} &= \arg \min \|\mathbf{Y} - \mathbf{L}_{(n)} - \mathbf{G}\|_F^2 + \varphi(\mathbf{G}), \\ \text{s.t. rank}(\mathbf{L}) &\leq r = \min(t, k), \end{aligned} \quad (16)$$

where  $\mathbf{G}$  is obtained by (15) when  $\mathbf{W} = \mathbf{Y} - \mathbf{L}$  is fixed. However, the GreB [35] strategy aims at dealing with the updating of  $\mathbf{L}$  in (16) and (3) to reduce failure caused by a biased estimate of the rank  $r$ .

In the GreB strategy, a bilateral factorization  $\mathbf{X} = \mathbf{U}\mathbf{V}$  is constructed to replace  $\mathbf{X}$ . In addition, the alternating optimization of  $\mathbf{U}$  and  $\mathbf{V}$  is obtained via QR decomposition.

$$\begin{cases} \mathbf{U}_{(i)} = \mathbf{Q}, & \mathbf{X}\mathbf{V}_{(i-1)}^T = \mathbf{Q}\mathbf{R} \\ \mathbf{V}_{(i)} = \mathbf{Q}^T\mathbf{X}, \end{cases} \quad (17)$$

where  $i$  is the number of the iteration. The detailed MCC-based LMD approach is summarized in Algorithm 1. The convergence of this approach has been proven to be linear [32].

**Algorithm 1** The MCC-Based LMD Algorithm

---

**Input:**  $Y \in \mathbb{R}^{M \times P}$ ,  $r, q, \sigma, \eta, \max = 100$   
**Output:**  $L = L^T$   
Initialize:  $G_{(0)} = 0, L_{(0)} = Y^T, n = 1$ ;  
Suppose a standard Gaussian matrix  $X_2 \in \mathbb{R}^{P \times r}$ ;  
**while** not converged **do**  
     $\hat{Y} = Y - G_{(n-1)}$ ;  
    **for**  $i = 1, \dots, q$  **do**  
         $X_1 = \hat{Y}X_2$ ;  
         $X_2 = \hat{Y}X_1$ ;  
    **end for**  
     $[Q, R] = \text{QR}(X_2, 0), X_2 = Q$ ;  
     $V = Q^T \hat{Y}, U_{(n)} = Q$ ;  
     $L_{(n)} = U_{(n)}V_{(n)} = QQ^T \hat{Y}, W_{(n)} = Y - L_{(n)}$ ;  
     $G = W_{(n)} - W_{(n)} \odot \kappa_\sigma(W_{(n)})$ ;  
    **if**  $\frac{\|L_{(n)} - L_{(n-1)}\|_F^2}{\|L_{(n-1)}\|_F^2} < \eta$  or  $n == \max$  **then**  
        **break**;  
    **end if**  
     $L \mathcal{D} L_{(n)}$ ;  
     $n = n + 1$ ;  
**end while**

---

## 3) THE ADAPTIVE ESTIMATE OF THE EFFECTIVE RANK VALUE

Since the energy of different frames is time-varying, using a fixed rank value in Algorithm 1 will reduce the recovery effect. Therefore, we introduce an energy threshold technique to update the effective rank value of the speech matrix  $L$ . Suppose that speech signal  $l(t)$  and additive noise  $s(t)$  are zero-mean Gaussian signals, i.e.,  $p(l) \sim \mathcal{N}(0, \sigma_l), p(s) \sim \mathcal{N}(0, \sigma_s)$ . It can be readily obtained that

$$E(y(t)y^H(t)) = E(l(t)l^H(t)) + E(s(t)s^H(t)), \quad (18)$$

$$\sigma_y = \sigma_l + \sigma_s, \quad (19)$$

where (19) is the simplified version of (18). In addition,  $\sigma_l$  and  $\sigma_s$  can be taken as the powers of the speech and noise, respectively. Let  $l_k(t)$  represents the signal synthesized by the first  $k$  largest eigenvalues, so we can obtain

$$\sigma(k) = \sigma_y - \sigma_1(k), \text{ s.t. } \sigma_y = \sigma_l + \sigma_s, \quad (20)$$

$$\sigma(k) - \sigma_s = \sigma_l - \sigma_1(k), \quad (21)$$

where  $\sigma(k)$  and  $\sigma_1(k)$  represent the difference power and the power of  $l_k(t)$ , respectively. However, in the non-Gaussian noise scene, formula (19) is not satisfied. The QR transformation can be used to obtain a unitary matrix from the noise matrix and whiten the noisy matrix to satisfy the above conditions. In addition, we use the formula (21) as the adaptive determination of the effective rank value of the speech matrix  $L$ . According to (21), when  $k$  is greater than the value of the actual rank  $r$  in Algorithm 1 (i.e.,  $\sigma(k) - \sigma_s < 0$ ), then this is a case of overestimating speech, because the recovered low-rank matrix  $L$  contains not only speech components but also partial noise components. In contrast, when  $k$  is less than the value of the actual rank  $r$  (i.e.,  $\sigma(k) - \sigma_s > 0$ ), it is a case of

underestimation where the speech signal has distortion. Thus, only the difference power in formula (21) is equal to or close to zero. The effective rank value of the speech matrix  $L$  can be obtained. More specifically, it can be given as

$$\tau = |\sigma(k) - \sigma_s|, \quad (22)$$

where the symbol  $\tau$  is a minimum value. Therefore, when the value of  $k$  satisfies the condition in (22), we can obtain the value of effective rank  $r$ . This approach of adaptively estimating the effective rank value is summarized in Algorithm 2.

**Algorithm 2** The Approach of Adaptively Estimating the Effective Rank Value

---

**Input:**  $Y \in \mathbb{R}^{M \times P}, \tau$   
**Output:**  $r = k$   
Initialize:  $k = 1$ ;  
Suppose  $y_1$  contains only the noise matrix  $S \in \mathbb{R}^{M \times P}$ ;  
**while** not converged **do**  
     $U \Lambda V_1^T = \text{SVD}(S)$ ;  
     $\sigma_s = \text{Sum}(\Lambda \Lambda^H)$ ;  
     $U \lambda V_2^T = \text{SVD}(Y)$ ;  
     $\sigma(k) = \text{Sum}(\lambda \lambda^H) - \sigma_1(k)$ ;  
    **if**  $\|\sigma(k) - \sigma_s\|_F^2 \leq \tau$  or  $k == P$  **then**  
        **break**;  
    **end if**  
     $k = k + 1$ ;  
**end while**

---

**C. EXPERIMENTAL SETTINGS**

For these experiments, we choose 30 different clean speech signals (sp01~sp30) based on the NOIZEUS database [43]. This database consists of 30 sentences with a sampling frequency of 8 kHz, and the sentences are produced by three female and three male. In addition, the additive noise included Gaussian white noise and different types of non-Gaussian noise (pink, babble, F16, and hfchannel) are taken from NOISEX-92 database [44]. The clean speech signal is corrupted by these different types of noise at 0, 5, 10, and 15 dB. The synthesized noisy signals are divided into frames of 32 ms each, and 50% frame overlap is used to ensure signal continuity. The window function is a 256-point Hamming window applied to suppress the Gibbs phenomenon. Overlap-add-synthesis is performed on the final enhanced speech signal to obtain the reconstructed speech matrix. Note that we perform a parallel recovery for all frames.

**D. PERFORMANCE METRICS**

We make use of three objective performance metrics to quantify and analyze the performance of these algorithms, namely, the segmental SNR (segSNR) [1], the perceptual evaluation of speech Quality (PESQ) [45], and the short-time objective intelligibility measure (STOI) [46], respectively. The segSNR metrics is served to reflect the suppression of interference

noise, defined as

$$\text{segSNR} = \frac{1}{N} \sum_{n=1}^N \left\{ 10 * \log_{10} \frac{\|I(n) - \hat{I}(n)\|^2}{\|I(n)\|^2} \right\}, \quad (23)$$

where  $I(n)$  and  $\hat{I}(n)$  denote the  $n$ -th frame of clean and reconstructed speech signal, respectively. The symbol  $N$  denotes the total number of speech signal frames. Moreover, The PESQ metric and STOI metric are commonly used to predict speech distortion and intelligibility, respectively. It is worth to note that they all calculated on the average.

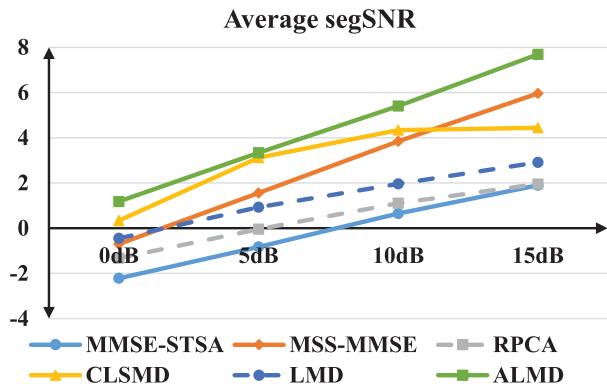
**E. EXPERIMENTAL RESULTS AND DISCUSSIONS**

1) PERFORMANCE IN GAUSSIAN WHITE NOISE

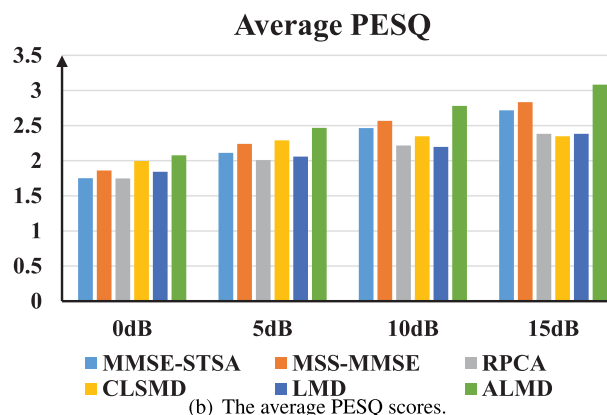
In this section, we will display the performance comparisons for the enhanced speech signal degraded by Gaussian white noise. The results are shown in Fig. 2.

As showing in Fig. 2(a), we give the relationship between the average segSNR and input SNR in the Gaussian white noise case. As expected, the above results show that both the average segSNR achieved by different methods are increasing with the input SNR improving. It also can be seen that the average segSNR generated by the ALMD approach is higher than that of the other methods (MMSE-STSA, Minimum mean square error estimator of magnitude squared spectrum (MSS-MMSE), RPCA, CLSMD, and LMD) across the SNR range of 0-15 dB. This may be attributed to the fact that the effective rank technique will be valuable for the suppression of noise. It’s interesting to note that the CLSMD solution will make very little noise attenuation when the SNR reaches > 10 dB. The MMSE-STSA method display the lower average segSNR than that of the other methods, since it is heavily sensitive to the correctness of VAD. Fig. 2(b) shows the results of the average PESQ scores obtained by different methods. The ALMD method has a significantly higher average PESQ scores than the baselines for the input SNR range of 0–15 dB with 5 dB steps. It also clearly displays that the average PESQ scores obtained by the MSS-MMSE increase rapidly with the SNR increases. This is duo to the fact that this method achieve satisfactory results by incorporating a *priori* SNR uncertainty [16]. In terms of average STOI scores, we can see from Fig. 2(c), the enhancement method presented herein can also yield significantly higher STOI scores than that of the other five algorithms. In contrast, LMD scheme has a very poor average STOI scores with the rank is fixed operation. These results manifest that the ALMD method outperforms the baselines in Gaussian white noise environment.

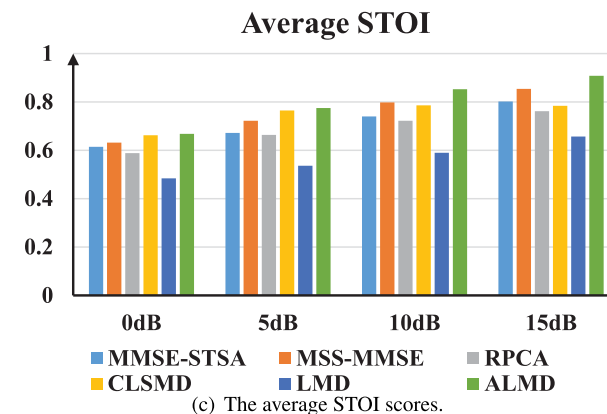
Fig. 3 shows the waveforms of reconstructed speech (“sp01”) obtained by different methods. The reconstruction is performed in a condition of very severe noise with additive Gaussian white noise at −5 dB. In this circumstance, it is easier to see the enhanced performance of the above mentioned enhancement schemes in the Gaussian white noise scene with a −5 dB SNR.



(a) The average segSNR.



(b) The average PESQ scores.



(c) The average STOI scores.

**FIGURE 2.** Performance comparisons of the different methods in the Gaussian white noise.

In Fig. 3, it is clearly shown that MMSE-STSA and RPCA led to a notable speech distortion in heavy white noise condition (−5 dB). We believe that this may be due to the following facts: MMSE-STSA needs accurate VAD detection and RPCA needs the prior knowledge of the signal source. Both of these conditions will be unavailable in strong noise conditions. Owing to the effective rank value estimate and the fact that a noisy phase approximation is not needed, the ALMD method yields significantly better waveform than the other methods (MSS-MMSE, CLSMD, and LMD). Unfortunately, there is some distortion. This may be because the strong noise (−5 dB) interferes with the estimate of the effective rank. Besides, the MSS-MMSE and CLSMD demonstrate

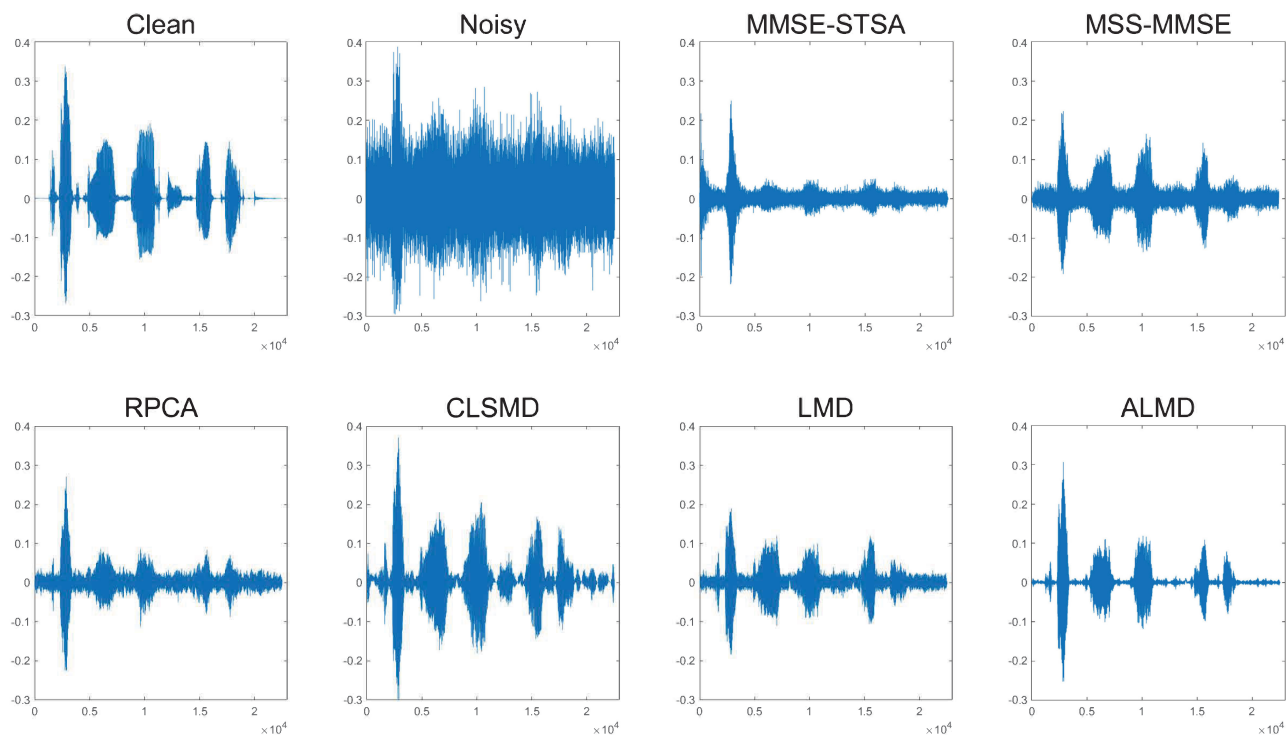


FIGURE 3. Time-domain waveform comparisons of speech (“sp01”) reconstruction in the white noise condition at  $-5$  dB SNR.

the similar noise suppression. Comparatively, the proposed method still has better performance in the white noise condition with a  $-5$  dB SNR.

## 2) PERFORMANCE IN NON-GAUSSIAN NOISE

Additionally, we will introduce the performance comparisons between our proposed algorithm and the baselines in several non-Gaussian noise scenarios (colored pink noise, non-stationary babble noise, F16 noise, and stationary hfchannel noise). First, we adopt the average segSNR metric to reflect the suppression of different types of non-Gaussian noise. The results are presented in Fig. 4.

Fig. 4 displays that the ALMD scheme has a substantially higher average segSNR than that of the other methods, as the SNR reaches  $>10$  dB, especially in the F16 noise condition. However, the average segSNR is comparable to the average segSNR obtained from MSS-MMSE approach at the lower SNR levels (0 dB, 5 dB). The CLSMD method achieves a lower average segSNR against these acquired by ALMD and MSS-MMSE for the SNR range of 0-10 dB in various noise conditions (pink, babble, and F16), but does with a relatively higher average segSNR improvement in hfchannel noise condition. This might be due to the hfchannel noise belongs to stationary noise. Consequently, hfchannel noise could be as a low rank component not like the highly non-stationary noise [47].

Second, the average PESQ scores are applied to reflect the speech distortion of the proposed ALMD approach for four different types of non-Gaussian noise. The results are shown in Fig. 5.

We could clearly see from Fig. 5 that our proposed method can obtain a substantially higher average PESQ scores in different kinds of non-Gaussian noise conditions with the input SNR range of 0–15 dB, except in the non-stationary babble noise scene at 0 dB and 5 dB SNR. However, the suggestion of this paper has similar effects to that of MSS-MMSE solution for the SNR less than 15 dB (i.e., 0 dB, 5 dB, and 10 dB). In comparison, the proposed method still perform well in different types of non-Gaussian noise conditions in terms of speech distortion. However, as mentioned earlier, under strong noise conditions, the estimate of the effective rank process will be affected. Thus, the enhanced speech will still contain a portion of the noise information. This fact can explain why the average PESQ scores obtained by the proposed scheme herein are sometimes lower than that obtained by the MSS-MMSE method at 0 dB.

Last, the average STOI scores are used to reflect the speech intelligibility of the presented approach in four different types of non-Gaussian noise conditions. The results are shown in Fig. 6.

We could clearly see from Fig. 6 that our presented approach achieves a significantly higher average scores in STOI metrics, and the improvement of STOI scores indicate that the proposed ALMD scheme can successfully recover the speech signal and that intelligibility is effectively improved. While several other reference algorithms cannot preserve the originality of reconstructed speech. The above results demonstrate that the ALMD scheme still could preserve the speech intelligibility in various types of non-Gaussian noise conditions, especially in low-SNR situations (0 dB, 5 dB).

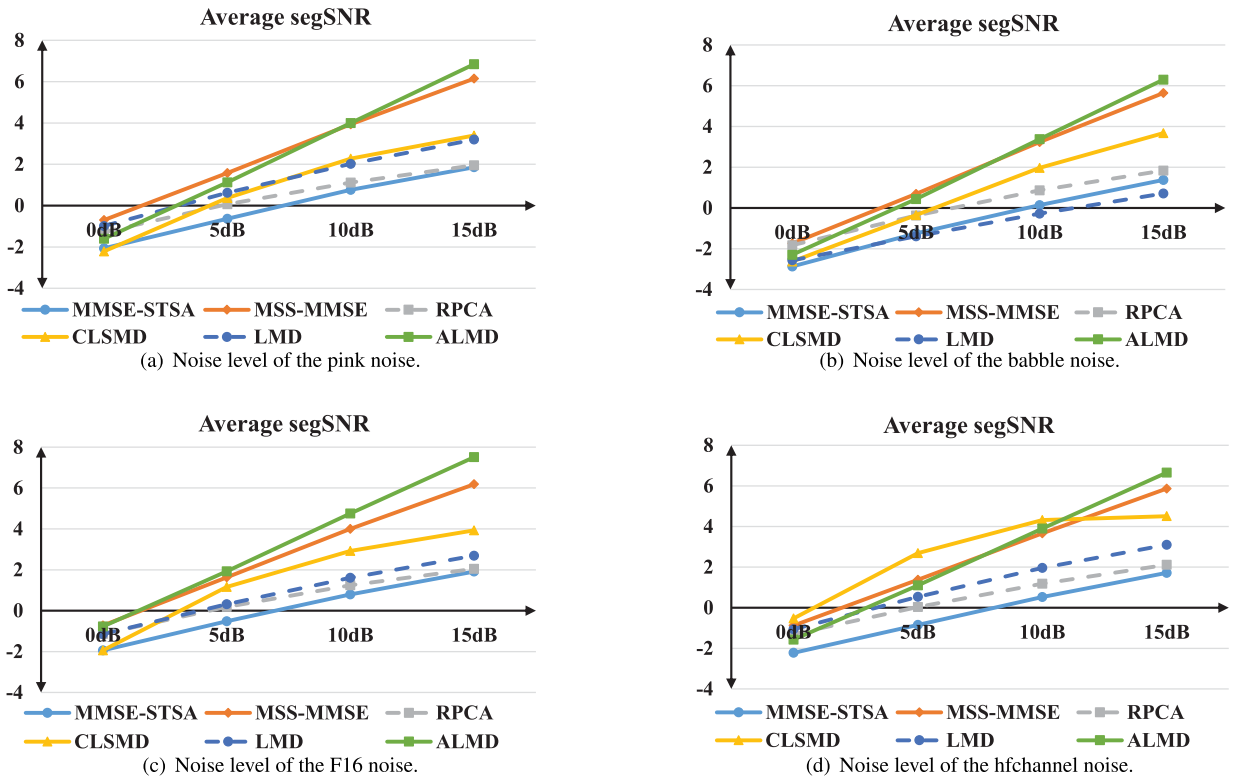


FIGURE 4. The average segSNR of the different methods in various types of non-Gaussian noise.

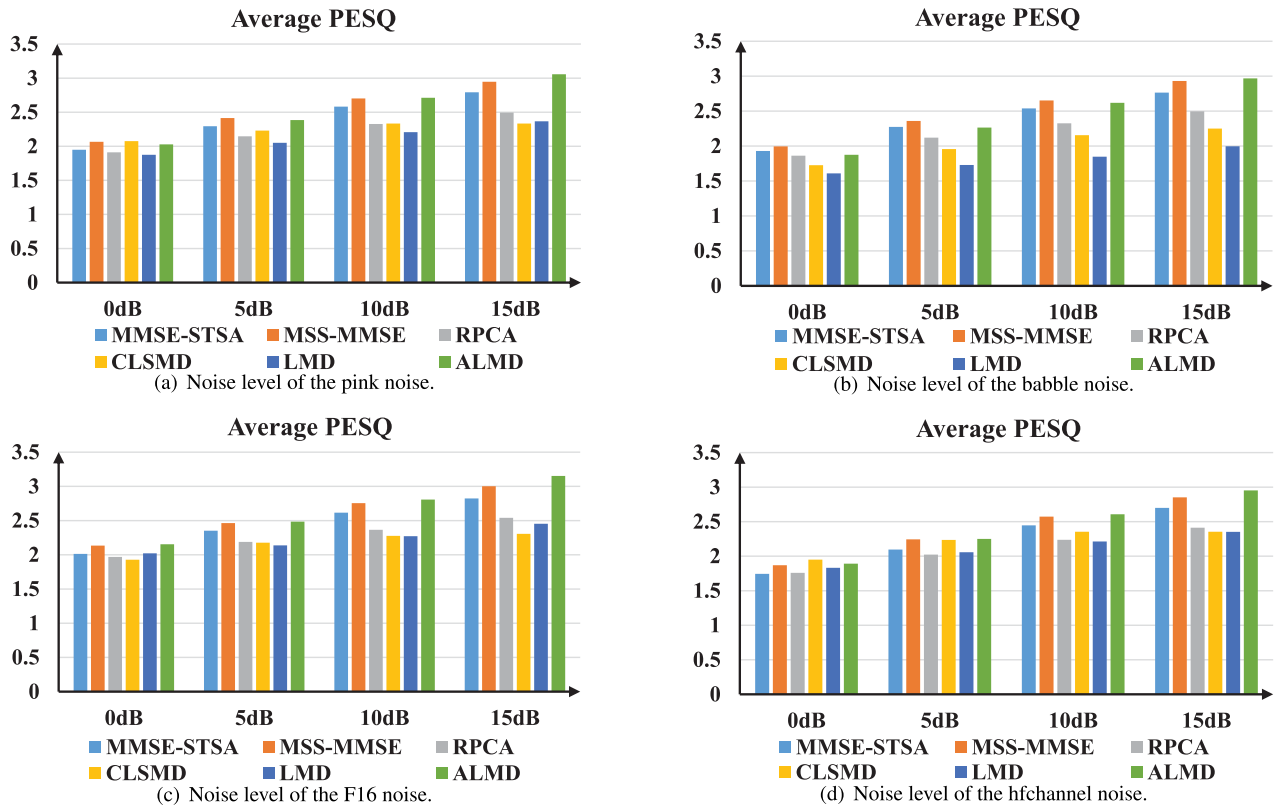


FIGURE 5. The average PESQ scores of the different methods in various types of non-Gaussian noise.



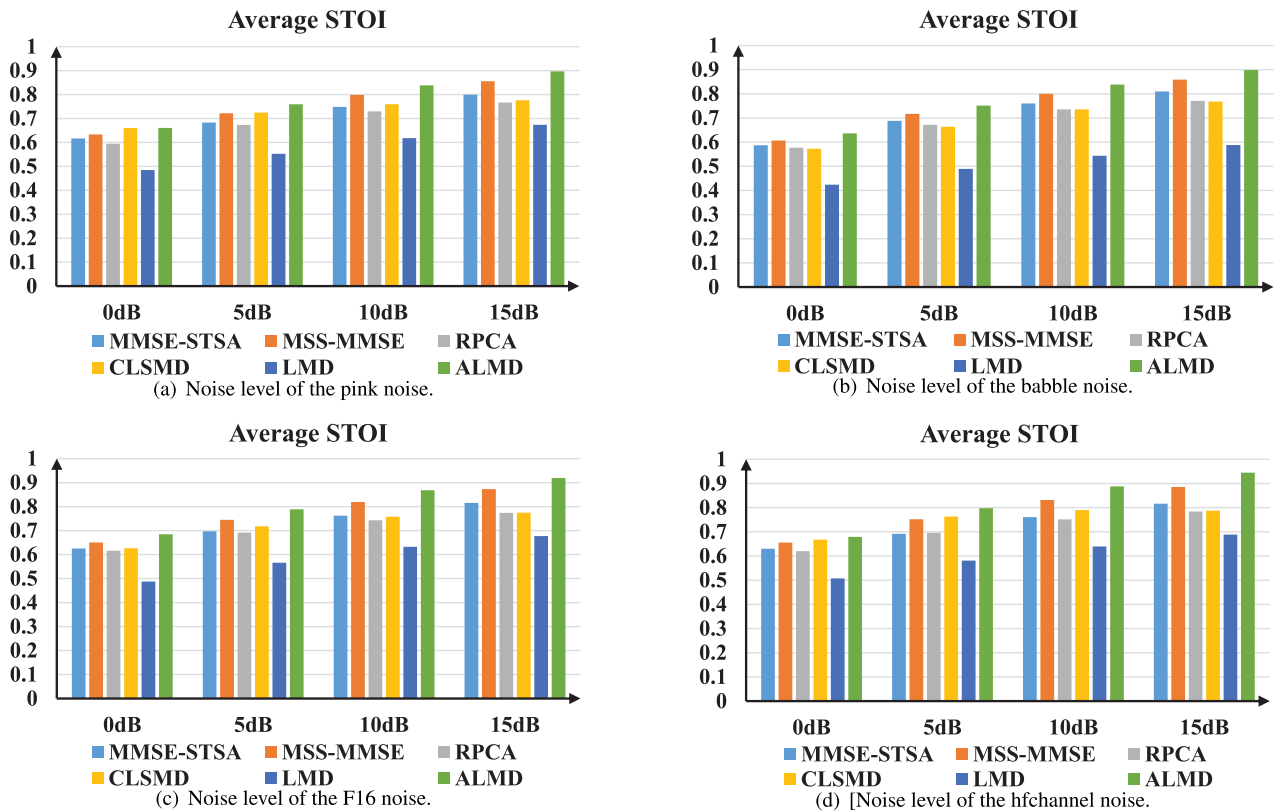


FIGURE 6. The average STOI scores of the different methods in various types of non-Gaussian noise.

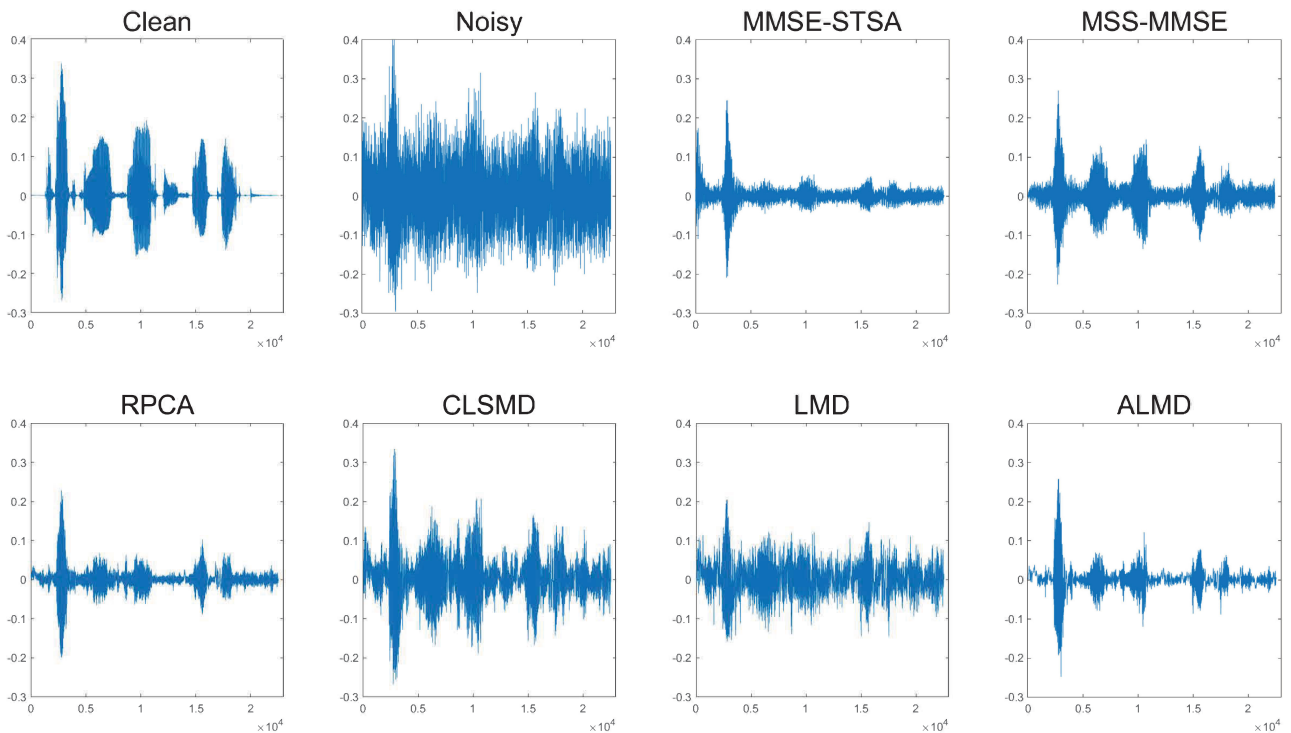


FIGURE 7. Time-domain waveform comparisons of speech (“sp01”) reconstruction in the F16 noise condition at  $-5$  dB SNR.

Fig. 7 shows the time-domain waveforms of reconstructed speech (“sp01”) obtained by six different enhancement

methods, which degraded by non-Gaussian F16 noise for input SNR at  $-5$  dB.

As can be seen clearly from Fig. 7, the MMSE-STSA, RPCA, CLSMD, and LMD methods have a disappointing loss of speech information. This is that the preconditions for the effective operation of these methods, such as noise obeys Gaussian distribution, prior signal knowledge, the rank of the noise is 1, and the value of the effective rank, which is no longer satisfied in strong non-Gaussian F16 noise condition (−5 dB). In addition, the time-domain waveform achieved by the MSS-MMSE is comparable with the performance of ALMD solution. However, it has some speech information lost in the high frequency part. Comparatively, in the case of non-Gaussian F16 noise, the ALMD method could achieve a better performance than the baselines at −5 dB. However, this method inevitably leaves much residual noise, which will be our future improvement.

#### IV. CONCLUSION

In this work, we present a novel speech enhancement algorithm based on ALMD scheme that exploits the fact that low-rank characteristics are shown by the speech signals in the time domain. The energy threshold technique is developed to adaptively update the effective rank value of each frame of the speech matrix. Additionally, the enhancement process in the time domain avoids phase loss, which is helpful to improve speech quality and intelligibility. The experimental results confirm that our improved algorithm outperformed the five baseline algorithms in terms of the segSNR, PESQ, and STOI, especially in the case of white noise. However, the proposed method also could yield significantly higher scores of PESQ and STOI than performed by the baselines in the non-Gaussian noise condition. Whilst it has low complexity without loss of evident speech intelligibility and quality. More significantly, our proposed method avoids the need for noise estimation, the VAD process, and approximation of the clean speech phase by the noisy phase.

#### REFERENCES

- [1] P. C. Loizou, *Speech Enhancement: Theory and Practice*. BocaRaton, FL, USA: CRC Press, 2013.
- [2] S. Markovich-Golan, S. Gannot, and W. Kellermann, "Combined LCMV-TRINICON beamforming for separating multiple speech sources in noisy and reverberant environments," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 2, pp. 320–332, Feb. 2017.
- [3] S. Y. Low, D. S. Pham, and S. Venkatesh, "Compressive speech enhancement," *Speech Commun.*, vol. 55, no. 6, pp. 757–768, Jul. 2013.
- [4] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 4, pp. 692–730, Apr. 2017.
- [5] O. Schwartz, S. Gannot, and E. A. P. Habets, "Multispeaker LCMV beamformer and postfilter for source separation and noise reduction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 5, pp. 940–951, May 2017.
- [6] S. Chakrabarty and E. A. P. Habets, "Time–frequency masking based online multi-channel speech enhancement with convolutional recurrent neural networks," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 4, pp. 787–799, Aug. 2019.
- [7] X. Li, L. Girin, S. Gannot, and R. Horaud, "Multichannel speech separation and enhancement using the convolutive transfer function," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 3, pp. 645–659, Mar. 2019.
- [8] D. Yee, H. Kamkar-Parsi, R. Martin, and H. Puder, "A noise reduction postfilter for binaurally linked single-microphone hearing aids utilizing a nearby external microphone," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 1, pp. 5–18, Jan. 2018.
- [9] J.-C. Wang, Y.-S. Lee, C.-H. Lin, S.-F. Wang, C.-H. Shih, and C.-H. Wu, "Compressive sensing-based speech enhancement," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 11, pp. 2122–2131, Nov. 2016.
- [10] K. Paliwal, K. Wójcicki, and B. Schwerin, "Single-channel speech enhancement using spectral subtraction in the short-time modulation domain," *Speech Commun.*, vol. 52, no. 5, pp. 450–475, May 2010.
- [11] C.-T. Lu, "Enhancement of single channel speech using perceptual-decision-directed approach," *Speech Commun.*, vol. 53, no. 4, pp. 495–507, Apr. 2011.
- [12] T. V. Sreenivas and P. Kinnappure, "Codebook constrained wiener filtering for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 4, no. 5, pp. 383–389, Sep. 1996.
- [13] I. Y. Soon and S. N. Koh, "Low distortion speech enhancement," *IEE Proc.-Vis., Image, Signal Process.*, vol. 147, no. 3, pp. 247–253, Jun. 2000.
- [14] Y. Wang and M. Brookes, "Speech enhancement using an MMSE spectral amplitude estimator based on a modulation domain Kalman filter with a gamma prior," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 5225–5229.
- [15] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 443–445, Apr. 1985.
- [16] Y. Lu and P. C. Loizou, "Estimators of the magnitude-squared spectrum and methods for incorporating SNR uncertainty," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 5, pp. 1123–1137, Jul. 2011.
- [17] A. Chinaev and R. Haeb-Umbach, "A generalized log-spectral amplitude estimator for single-channel speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 4980–4984.
- [18] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 4, pp. 251–266, Jul. 1995.
- [19] F. Asano, S. Hayamizu, T. Yamada, and S. Nakamura, "Speech enhancement based on the subspace method," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 497–507, Sep. 2000.
- [20] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 4, pp. 334–341, Jul. 2003.
- [21] Z. Chen and D. P. W. Ellis, "Speech enhancement by sparse, low-rank, and dictionary spectrogram decomposition," in *Proc. IEEE Workshop Appl. Signal Process. to Audio Acoust.*, New Paltz, NY, USA, Oct. 2013, pp. 1–4.
- [22] J. Huang, X. Zhang, Y. Zhang, X. Zou, and L. Zeng, "Speech denoising via low-rank and sparse matrix decomposition," *ETRI J.*, vol. 36, no. 1, pp. 167–170, Feb. 2014.
- [23] C. Sun, Q. Zhu, and M. Wan, "A novel speech enhancement method based on constrained low-rank and sparse matrix decomposition," *Speech Commun.*, vol. 60, pp. 44–55, May 2014.
- [24] P. Sun and J. Qin, "Low-rank and sparsity analysis applied to speech enhancement via online estimated dictionary," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1862–1866, Dec. 2016.
- [25] X. Li, M. Fan, L. Liu, and W. Li, "Distributed-microphones based in-vehicle speech enhancement via sparse and low-rank spectrogram decomposition," *Speech Commun.*, vol. 98, pp. 51–62, Apr. 2018.
- [26] D. Teng and D. Chu, "A fast frequent directions algorithm for low rank approximation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 6, pp. 1279–1293, Jun. 2019.
- [27] P. V. Giampouras, A. A. Rontogiannis, and K. D. Koutroumbas, "Alternating iteratively reweighted least squares minimization for low-rank matrix factorization," *IEEE Trans. Signal Process.*, vol. 67, no. 2, pp. 490–503, Jan. 2019.
- [28] J. Dong, Z. Xue, J. Guan, Z.-F. Han, and W. Wang, "Low rank matrix completion using truncated nuclear norm and sparse regularizer," *Signal Process., Image Commun.*, vol. 68, pp. 76–87, Oct. 2018.
- [29] S. Mavaddaty, S. M. Ahadi, and S. Seyedin, "A novel speech enhancement method by learnable sparse and low-rank decomposition and domain adaptation," *Speech Commun.*, vol. 76, pp. 42–60, Feb. 2016.
- [30] M. Hintermüller and T. Wu, "Robust principal component pursuit via inexact alternating minimization on matrix manifolds," *J. Math. Imag. Vis.*, vol. 51, no. 3, pp. 361–377, Aug. 2014.

- [31] N. Dionelis and M. Brookes, "Phase-aware single-channel speech enhancement with modulation-domain Kalman filtering," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 5, pp. 937–950, May 2018.
- [32] K. Guo, L. Liu, X. Xu, D. Xu, and D. Tao, "GoDec+: Fast and robust low-rank matrix decomposition based on maximum correntropy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2323–2336, Jun. 2018.
- [33] R. He, B.-G. Hu, W.-S. Zheng, and X.-W. Kong, "Robust principal component analysis based on maximum correntropy criterion," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1485–1494, Jun. 2011.
- [34] R. He, W.-S. Zheng, T. Tan, and Z. Sun, "Half-quadratic-based iterative minimization for robust sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 261–275, Feb. 2014.
- [35] T. Zhou and D. Tao, "Greedy bilateral sketch, completion & smoothing," *J. Mach. Learn. Res.*, vol. 31, pp. 650–658, Apr. 2013.
- [36] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, Jul. 2010.
- [37] C. Sun, Q. Zhang, J. Wang, and J. Xie, "Noise reduction based on robust principal component analysis," *J. Comput. Inf. Syst.*, vol. 10, no. 10, pp. 4403–4410, May 2014.
- [38] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Neural Inf. Process. Syst. (NIPS)*, Dec. 2009, pp. 2080–2088.
- [39] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2010, *arXiv:1009.5055*. [Online]. Available: <http://arxiv.org/abs/1009.5055>
- [40] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2117–2130, Sep. 2013.
- [41] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.
- [42] M. Nikolova and M. K. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery," *SIAM J. Sci. Comput.*, vol. 27, no. 3, pp. 937–966, Jan. 2005.
- [43] K. Wojcicki, M. Milacic, A. Stark, J. Lyons, and K. Paliwal, "Exploiting conjugate symmetry of the short-time Fourier spectrum for speech enhancement," *IEEE Signal Process. Lett.*, vol. 15, pp. 461–464, May 2008.
- [44] A. Varga and H. J. M. 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 Commun.*, vol. 12, no. 3, pp. 247–251, Jul. 1993.
- [45] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Salt Lake City, UT, USA, May 2001, pp. 749–752.
- [46] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of Time-Frequency weighted noisy speech," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011.
- [47] S. Y. Low and C. K. F. Yiu, "A study on the compressibility of speech for compressive speech enhancement," in *Proc. INTER-NOISE NOISE-CON Congr. Conf.*, Hong Kong, Aug. 2017, pp. 6186–6192.



**CHAO LI** received the B.S. degree in network engineering from the Zhengzhou University of Light Industry, Zhengzhou, China, in 2014, and the M.S. degree in signal and information processing from the Henan University of Technology, Zhengzhou, in 2017. He is currently pursuing the Ph.D. degree in information and communication engineering with the Beijing University of Posts and Telecommunications, Beijing, China. His research interests include acoustic signal processing and speech enhancement.



**TING JIANG** received the B.S., M.S., and Ph.D. degrees in communication and information system from Yanshan University, China, in 1982, 1988, and 2003, respectively. He is currently a Professor with the Beijing University of Posts and Telecommunications, Beijing, China. His research interests include wireless broadband interconnection, information theory, short distance wireless communication technological theory and application, and wireless sensor networks.



**SHENG WU** received the B.S. and M.S. degrees from the Beijing University of Posts and Telecommunications, Beijing, China, in 2004 and 2007, respectively, and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 2014. He was a Postdoctoral Researcher with the Tsinghua Space Center, Tsinghua University. He is currently an Associate Professor with the Beijing University of Posts and Telecommunications. He has published over 50 journal articles and conference papers. He also holds over 20 granted patents. His research interests are mainly in iterative signal processing, massive MIMO, and satellite communications. He has received the First Prize of the Science and Technology Award of the Chinese Institute of Electronics, in 2017, the Silver Medal of the 46th Geneva International Exhibition of Inventions, in 2018, and the Second Prize of the National Technological Invention Award of China, in 2018.



**JIANXIAO XIE** received the B.S. and M.S. degrees from the Beijing University of Posts and Telecommunications, in 2013 and 2016, respectively, where he is currently pursuing the Ph.D. degree in information and communication engineering. His research interests include MIMO and beamforming.

...