www.czasopisma.pan.pl PAN POLSKA AKADEMIA NAUK www.journals.pan.pl

PAN ARCHIVES OF ACOUSTICS
Vol. **44**, No. 4, pp. 681–692 (2019)
DOI: 10.24425/aoa.2019.129724



# Speech Enhancement Based on Constrained Low-rank Sparse Matrix Decomposition Integrated with Temporal Continuity Regularisation

Chengli SUN, Conglin YUAN*

*School of Information Engineering*
*Nanchang Hangkong University*
Nanchang, 330063, China
*Corresponding Author e-mail: yuan_conglin@163.com

(*received January 24, 2019; accepted July 22, 2019*)

Speech enhancement in strong noise condition is a challenging problem. Low-rank and sparse matrix decomposition (LSMD) theory has been applied to speech enhancement recently and good performance was obtained. Existing LSMD algorithms consider each frame as an individual observation. However, real-world speeches usually have a temporal structure, and their acoustic characteristics vary slowly as a function of time. In this paper, we propose a temporal continuity constrained low-rank sparse matrix decomposition (TCCLSMD) based speech enhancement method. In this method, speech separation is formulated as a TCCLSMD problem and temporal continuity constraints are imposed in the LSMD process. We develop an alternative optimisation algorithm for noisy spectrogram decomposition. By means of TCCLSMD, the recovery speech spectrogram is more consistent with the structure of the clean speech spectrogram, and it can lead to more stable and reasonable results than the existing LSMD algorithm. Experiments with various types of noises show the proposed algorithm can achieve a better performance than traditional speech enhancement algorithms, in terms of yielding less residual noise and lower speech distortion.

**Keywords:** speech enhancement; temporal continuity; low-rank and sparse decomposition.

## 1. Introduction

In real-world situations, speeches are generally mixed with various kinds of noise, such as background noise, channel noise, competing speakers sounds. Speech enhancement is an effective approach solving noise interference problems. Over the past several decades, people have developed a number of speech enhancement algorithms (Loizou, 2007). These algorithms can be classified into two categories, namely, unsupervised and supervised methods. Supervised methods require training phase to estimate parameters of the models considered for clean speech and noise using training data. Examples of these methods include hidden Markov model (HMM) based approaches (Mohammadiha, Arne, 2013), nonnegative matrix factorisation (NMF) based methods (Sun *et al.*, 2015), and deep neural network (DNN) based approaches (Kolbæk *et al.*, 2017). Compared with supervised methods, unsupervised methods do not need training stages and only require little noise segments to estimate the noise-related parameters for speech enhancement. Examples of these methods include spectral subtraction (Boll, 1979), Wiener filtering (WF) (Plapous *et al.*, 2006; Wiener, 1949), Minimum Mean Square Error (MMSE) estimation (Rugini, Banelli, 2016; Stark, Paliwal, 2011), subspace method (Ephraim, Van Trees, 1995; Hermus *et al.*, 2007; Hu, Loizou, 2003), and low-rank and sparse matrix decomposition (LSMD) based speech enhancement methods (Bando *et al.*, 2018; Li *et al.*, 2018; Sun *et al.*, 2014).

Spectral subtraction is a typical nonparametric method in the spectrum domain. This method assumes that additive noise and clean speech are independent of each other. Then a relatively clean speech signal can be obtained by subtracting the spectrum of the estimated noise from the spectrum of the noisy speech signal (Boll, 1979). Due to the inaccuracy of noise spectrum estimation, this approach is mainly plagued by musical noise problem (Lu, Loizou, 2008; Paliwal *et al.*, 2010). In contrast, MMSE spectral estimation is an effective method to solve the problem of excessive residual noise, and it has a better noise suppres-

sion at the low signal-to-noise-ratio (SNR) (Stark, Paliwal, 2011). Subspace method is another popular speech enhancement approach (Ephraim, Van Trees, 1995; Hermus *et al.*, 2007; Hu, Loizou, 2003). The principle of this method is to perform an orthogonal decomposition of the noisy observations into a signal subspace and a noise subspace. The critical step of subspace approach is splitting of two invariant subspaces associated with signal and noise via subspace decomposition, which can be achieved by Karhunen-Loeve transform or singular value decomposition (SVD). However, subspace decomposition is highly sensitive to the presence of strong noise, resulting in a large amount of residual noise within enhanced speech in low SNR situations.

In recent years, there has been an increasing interest in utilisation of LSMD for speech enhancement (Bando *et al.*, 2018; Li *et al.*, 2018). The main idea behind these kinds of speech enhancement methods is motivated by the robust principal component analysis (RPCA) theory (Cai *et al.*, 2010; Candes *et al.*, 2011; Xu *et al.*, 2012). RPCA was firstly applied in image processing field for background separation (Bouwmans *et al.*, 2017) and impulse noise removal (Jin, Ye, 2018). However, RPCA is unable to be modelled directly to solve the speech enhancement problem. In 2014, a constrained LSMD (CLSMD) based speech enhancement method was proposed (Sun *et al.*, 2014), which separates speech and noise spectrogram by setting constraints on rank and sparsity of each input audio frame. In this method, noise signal is regarded as a low-rank component in time-frequency (T-F) domain because noise spectra within different time frames are usually highly correlated with each other, while the speech signal is regarded as a sparse component since it is relatively sparse in T-F domain. CLSMD can obtain better performance in strong noise conditions, and does not need to know the exact distribution of noise signal. In 2018, Bayesian LSMD was proposed for multi-channel speech enhancement (Bando *et al.*, 2018), where multi-channel magnitude spectrogram can be decomposed into channel-wise low-rank noise spectrograms and sparse speech spectrogram common to all the channels. Then Sun *et al.* (2016) proposed a joint LSMD based subspace method for speech enhancement. In this method, LSMD is employed in time domain where low-rank component corresponds to enhanced speech and sparse component corresponds to noise signal. Kammi and Mollaei (2017) proposed a speech enhancement method with sparsity regularisation. In this method, speech enhancement is performed by minimising an appropriate objective function composed of a data fidelity term and sparsity imposing regularisation terms. Alternating direction method of multipliers is adapted to solve the objective function for speech enhancement.

In this paper, we propose a temporal continuity constrained low-rank sparse matrix decomposition (TCCLSMD) based speech enhancement method. In general, the TCCLSMD originated from CLSMD. Regrettably, the CLSMD method ignores the temporal continuity between adjacent speech frames in the process of speech enhancement, and it generates some isolated discrete points in the sparse matrix by LS decomposition. To solve these deficiencies and improve the quality of the speech system, the TCCLSMD-based method is proposed. Experiments with various types of noises show the proposed algorithm can achieve better performance than traditional speech enhancement algorithms, in terms of yielding less residual noise and lower speech distortion.

The rest of this paper is organised as follows. Section 2 simply reviews the related work. The algorithms of TCCLSMD are introduced in Sec. 3, and speech enhancement system is established later. The experimental results and some conclusions are explained in Secs 4 and 5, respectively.

## 2. Related work

### *2.1. Low-rank and sparse matrix decomposition*

The main idea behind LSMD-based speech enhancement method is motivated by the robust principal component analysis (RPCA) theory (Wright *et al.*, 2009). In the spectrum domain, the noise signal frames generally have a high degree of correlation, which mainly refers to the low-rank feature of the noise. Compared with noise signals, speech signals have a certain degree of sparsity, and they only are active at several frequency points. Some blind source separation algorithms are designed by the feature of speech sparseness, such as Sparse Component Analysis (SCA) (Shi, Song, 2016), and the sparse coding strategy (Zhen *et al.*, 2017).

Based on the assumption that the noise T-F matrix contains a low-rank structure and the speech T-F matrix contains a sparse structure, researchers expect to decompose the noisy speech by the RPCA method to obtain a sparse matrix corresponding to the speech signal and a low-rank matrix corresponding to the noise signal. The experimental results show that the obtained sparse matrix contains more noise interference, and the low-rank matrix contains more speech information. The main reason is that the speech spectrum matrix has a certain low-rank characteristics except the sparse characteristics. When the RPCA method is applied, the decomposition process is not constrained in advance. It will result in generating more residual noise information in the sparse matrix and more speech information in the low-rank matrix. The LS decomposition theory was applied to the speech enhancement problem (Mavaddaty *et al.*, 2016; Sun, Mu, 2015;

Sun *et al.*, 2016), and CLSMD algorithm was proposed to constrain the RPCA decomposition process.

For the T-F domain speech enhancement method, CLSMD states that speech T-F matrix and noise T-F matrix have sparse characteristics and low-rank characteristics, respectively. In the T-F domain, since noise signals within different time-frames are usually correlated with each other, the noise spectrum can be assumed to be in a low-rank subspace. On the other hand, speech signals can be regarded as relatively sparse (Sun *et al.*, 2014). Therefore, the enhanced speech can be obtained by the LS decomposition of noisy speech matrix.

Let us consider a mathematical model of speech enhancement based on CLSMD. Assuming that $s(t)$ is pure speech signal, $l(t)$ is an independent additive noise signal relative to $s(t)$, the noisy speech signal $y(t)$ is expressed as follows:

$$y(t) = s(t) + l(t) \tag{1}$$

by the short-time Fourier transform (STFT), $y(t)$ is transformed into

$$Y(n,k) = S(n,k) + L(n,k), \tag{2}$$

where $n = 1, ..., N$ and $k = 1, ..., K$ denote the frame and frequency indexes. Then perform complex conjugate operation on both sides of Eq. (2)

$$|Y(n,k)|^2 = |S(n,k)|^2 + |L(n,k)|^2 + 2|S(n,k)||L(n,k)|\cos(\Delta\theta), \tag{3}$$

where $\Delta\theta$ denotes the phase difference between $S(n,k)$ and $L(n,k)$. As mentioned earlier, $l(t)$ is an independent additive noise signal relative to $s(t)$, and the formula (3) can be simplified as

$$|Y(n,k)| \approx |S(n,k)| + |L(n,k)|. \tag{4}$$

According to the conclusion by Zhang and Zhao (2013), the formula (4) depends on two factors:

1) $\dfrac{|L(n,k)|}{|S(n,k)|} \to 0$ or $\dfrac{|L(n,k)|}{|S(n,k)|} \to \infty$;
2) $\cos(\Delta\theta) \to 1$.

In other words, when the SNR is much larger than 0 dB, and $\cos(\Delta\theta)$ is close to 1, the equal sign of the formula (4) will be established. This is the assumption of this paper, and the formula (4) can be simplified as follows:

$$\mathbf{Y} = \mathbf{S} + \mathbf{L}, \tag{5}$$

where $\mathbf{S}$ is the sparse matrix corresponding to pure speech, $\mathbf{L}$ is the low-rank matrix corresponding to noise. To recover the matrices $\mathbf{S}$ and $\mathbf{L}$, we can obtain the following optimisation formula:

$$\min_{L,S} \gamma\|\mathbf{S}\|_1 + \|\mathbf{L}\|_*, \qquad \text{s.t. } \mathbf{Y} = \mathbf{S} + \mathbf{L}, \tag{6}$$

where $\|\cdot\|_*$ is the nuclear norm, which is the sum of all singular values (Candes, Plan, 2010). $\|\cdot\|_1$ is the $l_1$-norm of a matrix and $\gamma$ is a balance parameter. Formula (6) is limited with rank and sparse constraints as follows:

$$\min_{L,S} \|\mathbf{Y} - \mathbf{S} - \mathbf{L}\|_F^2, \quad \text{s.t. } \operatorname{rank}(\mathbf{L}) \le r, \quad \operatorname{card}(\mathbf{S}) \le h, S_{ij} \ge 0, \tag{7}$$

where $\operatorname{card}(\cdot)$ represents the number of non-zero elements of a matrix, that is the $l_0$-norm of the matrix; $\|\cdot\|_F$ represents the Frobenius norm of a matrix; $r$ represents the rank constraint of the low-rank matrix $\mathbf{L}$, and $h$ represents the sparse constraint of the sparse matrix $\mathbf{S}$.

### *2.2. Sparse matrix reconstruction model*

The blind source separation is used for sparse matrix reconstruction. The so-called blind source separation refers to the method of extracting or separating each source signal obtained by the array receiving antenna or sensor, and the source signal without knowing a little prior knowledge (such as normality, independence, and stability). If there is no conditional constraint, this will be a multi-solution problem according to the traditional blind source separation method.

According to the unsupervised sound source separation (Abdali, NaserSharif, 2017; Virtanen, 2007), the amplitude spectrum of the speech frame is modelled by the following linear combination of basic functions:

$$s_t = \sum_{j=1}^{J} g_{j,t} b_j, \tag{8}$$

where $J$ is the number of basic functions, $g_{j,t}$ is the gain of the $j$-th basic function $b_j$, and $t$ is the frame sequence. By means of the formula (8), we can realise the reconstruction of the sparse matrix and obtain the corresponding matrix pattern as follows:

$$[S]_{k,t} = [B]_{k,j}[G]_{j,t}, \tag{9}$$

where $k = 1, ..., K$ and $t = 1, ..., T$ denote the frequency and frame indexes. It should be noted that the observation matrix $\mathbf{S}$ is the only known, and the elements of the basic matrix $\mathbf{B}$ and the gain matrix $\mathbf{G}$ to be estimated must be non-negative. The estimates of $\mathbf{B}$ and $\mathbf{G}$ can be obtained by minimising the cost function $c(\mathbf{B}, \mathbf{G})$, which is the weighted sum of the reconstruction error term $c_r(\mathbf{B}, \mathbf{G})$, the temporal continuity term $c_t(\mathbf{G})$, and the sparse term $c_s(\mathbf{G})$

$$c(\mathbf{B}, \mathbf{G}) = c_r(\mathbf{B}, \mathbf{G}) + \alpha c_t(\mathbf{G}) + \beta c_s(\mathbf{G}), \tag{10}$$

where $\alpha$ and $\beta$ are the weights of the last two items, respectively. In this paper, the temporal continuity term $c_t(\mathbf{G})$ is constructed by

$$c_t(\mathbf{G}) = \sum_{j=1}^{J} \frac{1}{\sigma_j^2} \sum_{t=2}^{T} (g_{t,j} - g_{t-1,j})^2, \tag{11}$$

where $g_{t,j}$ and $g_{t-1,j}$ denote the adjacent frames of the gain matrix, $\sigma_j$ is the $j$-th estimated standard deviation, and $\sigma_j = \sqrt{\frac{1}{T}\sum_{t=1}^{T} g_{t,j}^2}$. Regarding the sparse term $c_s(\mathbf{G})$, we can refer to the MAP source estimate (Kheder *et al.*, 2017) method and it is defined by

$$c_s(\mathbf{G}) = \sum_{j=1}^{J} \sum_{t=1}^{T} f\left(\frac{g_{t,j}}{\sigma_j}\right), \tag{12}$$

where $f(\cdot)$ selects the absolute value function $f(x) = |x|$. The initial values of the matrices $\mathbf{B}$ and $\mathbf{G}$ can be initialised by random positive values and then automatically updated by the multiplication update rule until the cost function is reduced to within the threshold or the number of iterations is greater than the set value. The update rule for $\mathbf{B}$ is given as follows:

$$\mathbf{B} \leftarrow \mathbf{B} \times \frac{\frac{\mathbf{S}}{\mathbf{BG}}\mathbf{G}^{\mathrm{T}}}{\mathbf{1}\mathbf{G}^{\mathrm{T}}}, \tag{13}$$

where $\mathbf{A} \times \mathbf{B}$ and $\mathbf{A}/\mathbf{B}$ are the multiplication and division of the corresponding elements of matrices $\mathbf{A}$ and $\mathbf{B}$, respectively. Then $\mathbf{1}$ is an all-one matrix with the same dimension as matrix $\mathbf{S}$. To obtain the update rule for $\mathbf{G}$, we need to establish the gradients of the reconstruction error term $c_r(\mathbf{B},\mathbf{G})$, the temporal continuity term $c_t(\mathbf{G})$, and the sparse term $c_s(\mathbf{G})$, which are defined as follows:

$$\nabla c_r(\mathbf{B},\mathbf{G}) = \mathbf{B}^{\mathrm{T}}\left(\mathbf{1} - \frac{\mathbf{S}}{\mathbf{BG}}\right), \tag{14}$$

$$[\nabla c_t(\mathbf{G})]_{j,t} = 2T\frac{g_{j,t} - g_{j,t+1}}{\sum_{i=1}^{T} g_{j,i}^2} - T\frac{2g_{j,t}\sum_{i=2}^{T}(g_{j,i} - g_{j,i-1})^2}{\left(\sum_{i=1}^{T} g_{j,i}^2\right)^2}, \tag{15}$$

$$[\nabla c_s(\mathbf{G})]_{j,t} = \frac{1}{\sqrt{\frac{1}{T}\sum_{i=1}^{T} g_{j,i}^2}} - \sqrt{T}\frac{g_{j,t}\sum_{i=1}^{T} g_{j,i}}{\left(\sum_{i=1}^{T} g_{j,i}^2\right)^{3/2}}, \tag{16}$$

the gradient of $c(\mathbf{B},\mathbf{G})$ is the weighted sum of $\nabla c_r(\mathbf{B},\mathbf{G})$, $\nabla c_t(\mathbf{B},\mathbf{G})$, and $\nabla c_s(\mathbf{B},\mathbf{G})$ as follows:

$$\nabla c(\mathbf{B},\mathbf{G}) = \nabla c_r(\mathbf{B},\mathbf{G}) + \alpha\nabla c_t(\mathbf{G}) + \beta\nabla c_s(\mathbf{G}) \tag{17}$$

which is rewritten as a subtraction form $\nabla c(\mathbf{B},\mathbf{G}) = \nabla c^{+}(\mathbf{B},\mathbf{G}) - \nabla c^{-}(\mathbf{B},\mathbf{G})$, where

$$\nabla c^{-}(\mathbf{B},\mathbf{G}) = \nabla c_r^{-}(\mathbf{B},\mathbf{G}) + \alpha\nabla c_t^{-}(\mathbf{G}) + \beta\nabla c_s^{-}(\mathbf{G}), \tag{18}$$

$$\nabla c^{+}(\mathbf{B},\mathbf{G}) = \nabla c_r^{+}(\mathbf{B},\mathbf{G}) + \alpha\nabla c_t^{+}(\mathbf{G}) + \beta\nabla c_s^{+}(\mathbf{G}), \tag{19}$$

where

$$\nabla c_r^{-}(\mathbf{B},\mathbf{G}) = \frac{\mathbf{B}^{\mathrm{T}}\mathbf{S}}{\mathbf{BG}},$$

$$\nabla c_r^{+}(\mathbf{B},\mathbf{G}) = \mathbf{B}^{\mathrm{T}},$$

$$\nabla c_t^{-}(\mathbf{B},\mathbf{G}) = T\frac{2g_{j,t}\sum_{i=2}^{T}(g_{j,i} - g_{j,i-1})^2}{\left(\sum_{i=1}^{T} g_{j,i}^2\right)^2},$$

$$\nabla c_t^{+}(\mathbf{B},\mathbf{G}) = 2T\frac{g_{j,t} - g_{j,t+1}}{\sum_{i=1}^{T} g_{j,i}^2},$$

$$\nabla c_s^{-}(\mathbf{B},\mathbf{G}) = \sqrt{T}\frac{g_{j,t}\sum_{i=1}^{T} g_{j,i}}{\left(\sum_{i=1}^{T} g_{j,i}^2\right)^{3/2}},$$

$$\nabla c_s^{+}(\mathbf{B},\mathbf{G}) = \frac{1}{\sqrt{\frac{1}{T}\sum_{i=1}^{T} g_{j,i}^2}}.$$

That is to say, $\nabla c_r^{-}(\mathbf{B},\mathbf{G})$, $\nabla c_t^{-}(\mathbf{B},\mathbf{G})$, and $\nabla c_s^{-}(\mathbf{B},\mathbf{G})$ represent the subtractions in Eqs (14), (15) and (16), respectively, and $\nabla c_r^{+}(\mathbf{B},\mathbf{G})$, $\nabla c_t^{+}(\mathbf{B},\mathbf{G})$, and $\nabla c_s^{+}(\mathbf{B},\mathbf{G})$ represent the subtracted numbers in (14), (15), and (16), respectively.

Finally, we get the update rule for $\mathbf{G}$ as follows:

$$\mathbf{G} \leftarrow \mathbf{G} \times \frac{\nabla c^{-}(\mathbf{B},\mathbf{G})}{\nabla c^{+}(\mathbf{B},\mathbf{G})}. \tag{20}$$

## 3. TCCLSMD-based speech enhancement method

The speech signal generally exhibits short-term stability, and the data between the adjacent speech frames obtained by framing the time-lapse sequence using the window function has continuity. When the LS decomposition theory is applied to extract the sparse components through the hard threshold function, since the temporal continuity characteristics of the speech are not taken into consideration, it will generate some isolated discrete points in the sparse matrix. In view of this drawback, the LS decomposition is constrained by introducing the temporal continuity of speech (Virtanen, 2007); adding weight adjustment parameters in the process of constructing mathematical model; reconstructing the extracted sparse components; and reducing isolated discrete points to make sparse matrices more consistent with the speech spectrum distribution.

### 3.1. Optimisation algorithm for TCCLSMD

**Problem 1 (TCCLSMD).** Suppose a noisy signal matrix is given as $\mathbf{Y} = \mathbf{S} + \mathbf{L} + \mathbf{R}_E$, where $\mathbf{S}$ is the sparse matrix corresponding to pure speech, $\mathbf{L}$ is the low-rank matrix corresponding to noise, and $\mathbf{R}_E$ is the reconstruction error matrix. Suppose $r$ represents the rank constraint of the low-rank matrix $\mathbf{L}$, and $h$ represents the sparse constraint of the sparse matrix $\mathbf{S}$. To recover the matrices $\mathbf{S}$ and $\mathbf{L}$, we refer to formula (7) to get

$$\begin{aligned} &\min_{L,S} \|\mathbf{Y} - \mathbf{S} - \mathbf{L}\|_F^2, \\ &\text{s.t. } \operatorname{rank}(L) \le r, \qquad \operatorname{card}(S) \le h, S_{ij} \ge 0, \end{aligned} \tag{21}$$

where $\operatorname{card}(\cdot)$ represents the number of non-zero elements of a matrix, that is the $l_0$-norm of a matrix; $\|\cdot\|_F$ represents the Frobenius norm of a matrix. To solve for $\mathbf{S}$ and $\mathbf{L}$, we transform the problem into two sub-problems as follows (Liu, Peng, 2018):

$$\begin{aligned} \mathbf{L}_t &= \arg \min_{\operatorname{rank}(L) \le r} \|\mathbf{Y} - \mathbf{L} - \mathbf{S}_{t-1}\|_F^2, \\ \mathbf{S}_t &= \arg \min_{|S|_0 \le h, S_{ij} \ge 0} \|\mathbf{Y} - \mathbf{L}_t - \mathbf{S}\|_F^2. \end{aligned} \tag{22}$$

To solve the fixed rank problem of $(22)_1$ by the SVD-based method (Cai *et al.*, 2010), it is assumed that the singular value of the matrix $\mathbf{Y}$ is decomposed into

$$\operatorname{SVD}(\mathbf{Y}) = \mathbf{U}\sum\mathbf{V}^{\mathrm{T}}, \tag{23}$$

where $\mathbf{U}$ and $\mathbf{V}$ are the left and right feature vectors of $\mathbf{Y}$, respectively, and $\sum$ is the diagonal matrix whose diagonal is the eigenvalue. Then the low-rank matrix $\mathbf{L}$ is solved as follows:

$$L_i = \sum_i^r \lambda_i U_i V_i^{\mathrm{T}}, \tag{24}$$

where $\lambda_i$ is the singular value of $\sum$, and $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_{r-1} \ge \lambda_r$. Since $(22)_2$ is a non-convex function, it is unable to be solved by the optimisation theory. Then the sparse matrix $\mathbf{S}$, can be estimated by introducing a hard threshold function (Tan *et al.*, 2013), which is defined as follows:

$$\mathbf{S}_t = (\mathbf{Y} - \mathbf{L}_t) \otimes [(\mathbf{Y} - \mathbf{L}_t) > \mathbf{T}], \tag{25}$$

where $\otimes$ denotes the element-wise multiplication, and $\mathbf{T} > 0$. It is worth noting that $\mathbf{S}_t$ obtained by (25) is nonnegative. To reconstruct $\mathbf{S}_t$ by the sparse matrix reconstruction model, we use (13) and (20) to obtain

$$\mathbf{S}_t = \mathbf{B}_t \cdot \mathbf{G}_t, \tag{26}$$

where, $\mathbf{B}_t$ and $\mathbf{G}_t$ denotes the basic matrix and the gain matrix, respectively. Note that $\mathbf{B}_t$ and $\mathbf{G}_t$ are limited to be nonnegative. Then a solution model of TCCLSMD can be established as follows:

$$\begin{aligned} \mathbf{L}_t &= \mathbf{U}_{n\times r}\sum\nolimits_{r\times r}\mathbf{V}_{r\times k}^{\mathrm{T}}, \\ \operatorname{SVD}(\mathbf{Y} - \mathbf{S}_{t-1}) &= \mathbf{U}\sum\mathbf{V}^{\mathrm{T}}, \\ \mathbf{S}_t &= \mathbf{B}_t \cdot \mathbf{G}_t \end{aligned} \tag{27}$$

Then, we can obtain the following optimisation algorithm for TCCLSMD in Algorithm 1. Subalgorithm of reconstruction of the sparse matrix can be seen in Algorithm 2.

**Algorithm 1.**
The optimisation algorithm of TCCLSMD.

Given $r$, $T$, $t_{\max}$, $\mu$;
Initialize: $Y_0 = Y$, $S_t = [0]_{N\times K}$, $t = 0$
While not converged do
$\mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}} = \operatorname{SVD}(\mathbf{Y}_t)$;
$\mathbf{L}_t = \sum_{i=0}^{r} \lambda_i U_i V_i^{\mathrm{T}}$;
%update of sparse matrix $\mathbf{S}$
$\mathbf{X}_t = \mathbf{Y}_t - \mathbf{L}_t + \mathbf{S}_t$;
$\mathbf{S}_t = \mathbf{X}_t \otimes (\mathbf{X}_t > \mathbf{T})$;
$\mathbf{S}_t = \mathbf{B}_t\mathbf{G}_t$ (Algorithm 2);
If $\|\mathbf{Y} - \mathbf{L}_t - \mathbf{S}\|_F^2 / \|\mathbf{Y}\|_F^2 \le \mu$ or $t == t_{\max}$
break;
end
$\mathbf{Y}_t = \mathbf{L}_t + \mathbf{X}_t - \mathbf{S}_t$;
$t = t + 1$;
end while
output: $\mathbf{L} = \mathbf{L}_t$, $\mathbf{S} = \mathbf{S}_t$

**Algorithm 2.**
Reconstruction of sparse matrix.

Input: $M$, $\varepsilon$, $\alpha$, $\beta$, max;
Initialize:
$t = 1$, $\mathbf{B}_0 = \operatorname{randn}(\mathbf{K}, \mathbf{M})$, $\mathbf{G}_0 = \operatorname{randn}(\mathbf{M}, \mathbf{J})$;
While not converged do
% update of basic matrix $\mathbf{B}$
$\mathbf{B}_t = \mathbf{B}_{t-1} \times \dfrac{\frac{\mathbf{S}}{\mathbf{B}_{t-1}\mathbf{G}_{t-1}}\mathbf{G}_{t-1}^{\mathrm{T}}}{\mathbf{1}\mathbf{G}_{t-1}^{\mathrm{T}}}$;
% update gain matrix $\mathbf{G}$
$\mathbf{G}_t = \mathbf{G}_{t-1} \times \dfrac{\nabla c^-(\mathbf{B}_{t-1}, \mathbf{G}_{t-1})}{\nabla c^+(\mathbf{B}_{t-1}, \mathbf{G}_{t-1})}$;
If $c(\mathbf{B}, \mathbf{G}) \le \varepsilon$ or $t == \max$
break;
end
$t = t + 1$;
end while
output: $\mathbf{B} = \mathbf{B}_t$, $\mathbf{G} = \mathbf{G}_t$, $\mathbf{S} = \mathbf{B}\mathbf{G}$.

### 3.2. TCCLSMD-based speech enhancement method

Based on the aforementioned algorithm, we utilise the analysis-modification-synthesis (AMS) framework (Paliwal *et al.*, 2010) to build up TCCLSMD-based speech enhancement system. Figure 1 shows the block diagram of the TCCLSMD-based speech enhancement method. First, the amplitude spectrum of the noisy speech signal is framed by a window function, and each frame data is used as a matrix column, then the frame data is transformed into a frequency domain structure matrix by a STFT.

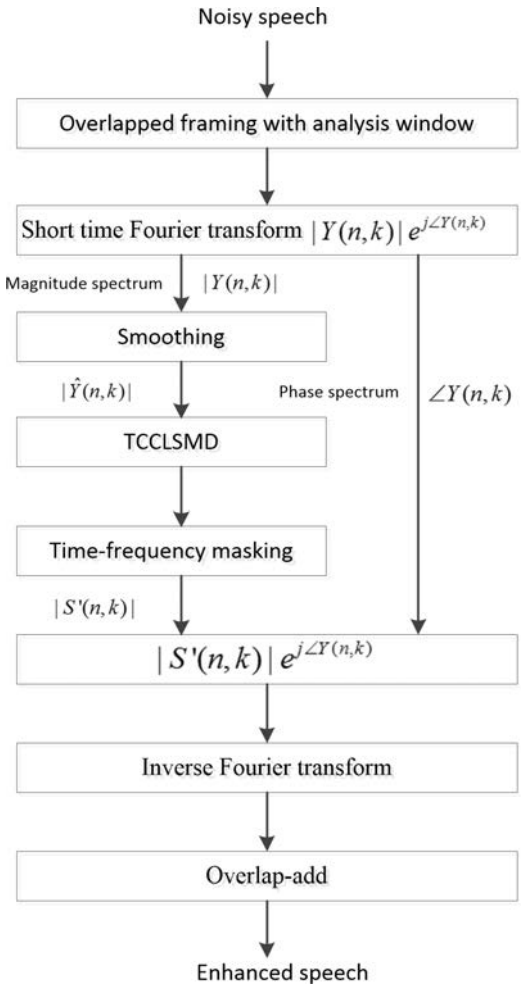


Fig. 1. Block diagram of the TCCLSMD-based speech enhancement method.

Secondly, a three-point median filter is used to smooth the spectral magnitude of noisy signal:

$$|\widehat{Y}(n,k)| = (|Y(n-1,k)| + |Y(n,k)| + |Y(n+1,k)|)/3. \tag{28}$$

Since the obtained $|\widehat{Y}(n,k)|$ is going to be stacked as columns to recover the noisy matrix $\mathbf{Y}$ in the T-F domain, then $\mathbf{S}$ and $\mathbf{L}$ are restored from $\mathbf{Y}$ by the TCCLSMD algorithm. Moreover, the T-F masking process needs to be performed. It is worth noting that phase spectrum has little effect on the enhanced speech (Shannon, Paliwal, 2006). So we can obtain the enhanced speech spectrum as follows:

$$\widehat{S}(n,k) = |S'(n,k)|e^{j\angle Y(n,k)}. \tag{29}$$

To recover the enhanced speech, we make use of the inverse Fourier transform and least-squares overlap add synthesis (Quatieri, 2002) in the last step.

## 4. Experiments

To compare the proposed method with mainstream speech enhancement methods, especially the CLSMD method, we take 30 sentences (sp1 ~ sp30) from the NOIZEUS database (Hu, Loizou, 2008). The sentences were affected by several types of noises, such as street, car, exhibition, train, station, white, hfchannel, pink, and F16 at 0 dB, 5 dB, 10 dB, and 15 dB, respectively. In addition, the sentences were all sampled at 8 kHz. In this TCCLSMD method, a Hamming window of 300 sample points (37.5 ms) with 40% frame overlap is used to segment the input signal into frames, and the STFT points is 1024 in the frequency domain. For the evaluation measures, the segment SNR measure and Perceptual Evaluation of Speech Quality (PESQ) measure are proposed.

### 4.1. Effects of temporal continuity and sparse weight on performance

Firstly, we check the influence of temporal continuity weight on the speech enhancement performance. We used sentences sp1 ~ sp30 as clean speech data and white noise as interference noise for every sentence. All segSNR and PESQ scores were averaged over the 30 test sentences. To eliminate the influence of sparse weight on the selection of temporal continuity weight $\alpha$, $\beta$ is first set to zero in this experiment, and then the segSNR and PESQ scores of five different interval points of 0.001 to 10 are tested respectively.

Figure 2a shows the segSNR and PESQ score line graphs for different $\alpha$ in a white noise environment. It can be seen from the figure that with the gradual increase of weights, the segSNR and PESQ scores of the white noise with different SNRs show a trend of rising first and then decreasing. When $\alpha$ is taken as 0.01, the best experimental results are obtained. Therefore, the subsequent experiments will select 0.01 as the weight.

In this case, we tested the influence of different $\beta$ on this experiment under white noise environment. Figure 2b shows the scores of segSNR and PESQ corresponding to different $\beta$. It can be found that the segSNR and PESQ scores tend to be stable, so the effect of sparse weight $\beta$ on the experimental results has little effect. This paper selects $\beta = 1$ as the weight.

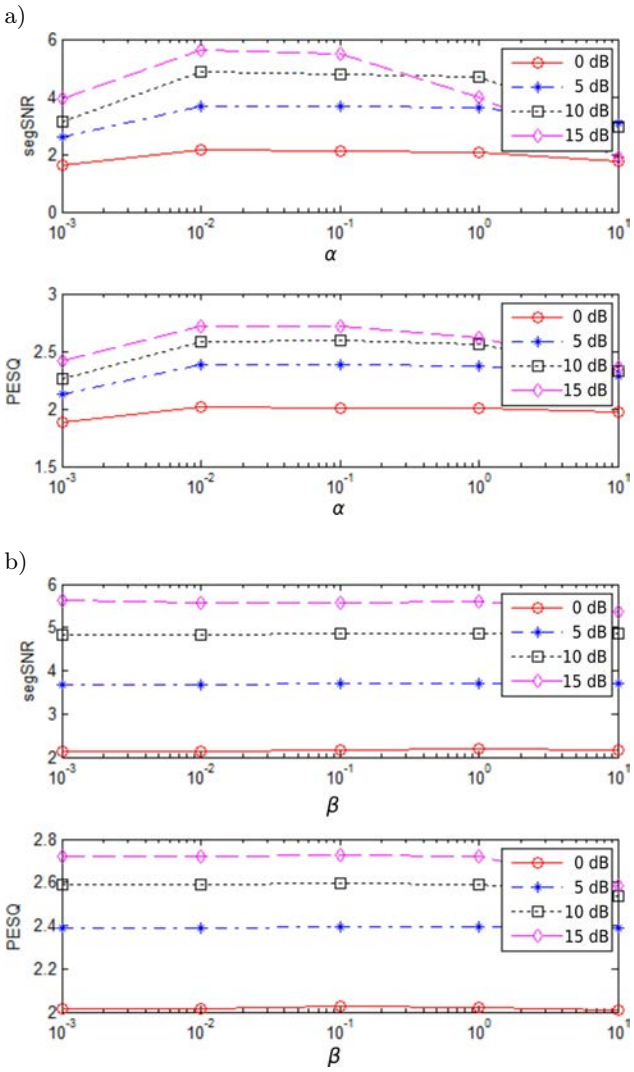


Fig. 2. Line graphs of segSNR and PESQ scores for different $\alpha$ and $\beta$ at white noise: a) the scores for $\alpha$, b) the scores for $\beta$.

### *4.2. Performance comparison with other speech enhancement methods*

The proposed method is going to be compared with mainstream speech enhancement approaches which include several typical approaches: spectral subtraction (Boll, 1979), subspace method (Moor, 1993), WF algorithm (Scalart, Vieira-Filho, 1996), minimum mean-square error (MMSE) method (Cohen, 2004), and a popular method, maximum a posterior estimator of magnitude-squared spectrum (MSS_MAP) algorithm (Paliwal *et al.*, 2012). More importantly, since the proposed method is optimised based on the CLSMD method (Sun *et al.*, 2014), the focus of this paper is compared with the CLSMD method.

For the TCCLSMD method, the coefficient of rank constraint $r$ was set to 1, iteration number was set to $t_{\max} = 100$, the coefficient of error control was set to $\mu = 0.001$; in the sub-algorithm, reconstruction iteration number was set to $M = 300$, the coefficient of error range precision control was set to $\varepsilon = 0.0001$, the weight of temporal continuity was set to $\alpha = 0.01$, and sparse weight was set to $\beta = 1$. For all related methods, we use MATLAB software to get experimental results (Sun *et al.*, 2014).

Table 1 shows the segSNR and PESQ scores for the 0 dB, 5 dB SNR noisy speech processed by the TCCLSMD method and the CLSMD method. It can be seen from Table 1 that the TCCLSMD method has a significantly better speech enhancement effect in the strong noise environment than the CLSMD method, except for the PESQ score of the 0 dB SNR in exhibition and F16 noise cases, which is slightly lower than the CLSMD method, respectively.

Tables 2 and 3 show respectively the comparison of segSNR and PESQ scores for seven speech enhancement methods with different SNRs. From the aspect of segSNR measure, the score of TCCLSMD method is slightly lower than CLSMD method under the street noise at 0 dB SNR; it has the highest score in other strong noise environments, indicating the suppression ability of TCCLSMD method has a good advantage over other methods at strong noise. Compared with the CLSMD method, the scores of the street noise at 0 dB SNR and the pink noise at 5 dB SNR are slightly lower than that of the CLSMD method. In other cases, the score is higher than the CLSMD method. By the PESQ measure, it can be found that the highest PESQ scores are obtained under most noises of 0 dB and 5 dB SNR in this proposed method.

Table 1. TCCLSMD and CLSMD methods for segSNR and PESQ scores for 0 dB, 5 dB SNR noisy speeches.

| Measure | Noise | TCCLSMD (0 dB) | CLSMD (0 dB) | TCCLSMD (5 dB) | CLSMD (5 dB) |
|---|---|---|---|---|---|
| segSNR | F16 | **–0.33** | –0.40 | **1.66** | 1.51 |
| | volvo | **0.61** | 0.56 | **–0.26** | –0.37 |
| | exhibition | **–1.72** | –1.73 | **0.37** | 0.36 |
| | station | **–1.58** | –1.58 | **–1.58** | –1.58 |
| PESQ | F16 | **1.98** | 1.98 | **2.33** | 2.31 |
| | volvo | **2.70** | 2.68 | **2.82** | 2.80 |
| | exhibition | **1.45** | 1.45 | **1.94** | 1.93 |
| | station | **1.66** | 1.65 | **2.11** | 2.10 |

Table 2. Comparison of segSNR scores for seven voice enhancement methods for various noisy speeches.

| SNR [dB] | Noise | TCCLSMD | CLSMD | MSS | SS | Subspace | MMSE | Wiener96 |
|---|---|---|---|---|---|---|---|---|
| 0 | White | **2.18** | 2.15 | 0.36 | −1.00 | 0.47 | −0.76 | −0.46 |
| | Hfchannel | **0.73** | 0.71 | 0.06 | −1.00 | −1.49 | −0.88 | −0.43 |
| | Pink | **0.17** | 0.09 | 0.07 | −0.70 | −3.23 | −0.70 | −0.22 |
| | Train | **−0.16** | −0.18 | −1.58 | −1.45 | −2.45 | −2.49 | −1.06 |
| | Street | **−1.39** | −1.39 | −1.71 | −0.98 | −2.62 | −2.07 | −0.84 |
| | Car | **−0.43** | −0.51 | −0.79 | −0.73 | −2.40 | −1.29 | −0.29 |
| 5 | White | **3.71** | 3.69 | 2.63 | 0.60 | 3.00 | 0.47 | 0.79 |
| | Hfchannel | **2.62** | 2.60 | 2.31 | 0.95 | 1.12 | 0.37 | 0.83 |
| | Pink | **2.00** | 1.73 | 2.44 | 1.20 | −0.61 | 0.63 | 1.07 |
| | Train | **1.68** | 1.65 | 0.93 | 0.86 | 0.16 | −0.70 | 0.42 |
| | Street | **0.92** | 0.92 | 0.54 | 0.71 | −0.22 | −0.80 | 0.22 |
| | Car | **1.43** | 1.30 | 1.31 | 1.19 | −0.20 | 0.08 | 0.96 |
| 10 | White | **4.87** | 4.86 | 5.14 | 3.14 | 5.53 | 1.51 | 2.12 |
| | Hfchannel | **4.22** | 4.22 | 4.80 | 3.32 | 3.78 | 1.85 | 2.26 |
| | Pink | **3.40** | 3.05 | 5.09 | 3.48 | 2.21 | 1.97 | 2.39 |
| | Train | **3.40** | 3.36 | 3.59 | 3.12 | 2.73 | 0.87 | 1.91 |
| | Street | **2.56** | 2.55 | 3.51 | 3.26 | 2.37 | 0.89 | 1.80 |
| | Car | **3.27** | 3.12 | 3.89 | 3.36 | 2.44 | 1.56 | 2.36 |
| 15 | White | **5.61** | 5.64 | 7.83 | 5.30 | 8.00 | 3.10 | 3.37 |
| | Hfchannel | **5.25** | 5.24 | 7.54 | 5.44 | 6.42 | 2.99 | 3.30 |
| | Pink | **4.32** | 4.02 | 8.03 | 5.63 | 5.06 | 3.13 | 3.35 |
| | Train | **4.59** | 4.47 | 6.73 | 5.38 | 5.36 | 2.34 | 2.94 |
| | Street | **3.70** | 3.68 | 6.25 | 5.07 | 5.29 | 2.22 | 2.80 |
| | Car | **4.42** | 4.31 | 6.84 | 5.76 | 5.42 | 2.81 | 3.40 |

Table 3. Comparison of PESQ scores for seven voice enhancement methods for various noisy speeches.

| SNR [dB] | Noise | TCCLSMD | CLSMD | MSS | SS | Subspace | MMSE | Wiener96 |
|---|---|---|---|---|---|---|---|---|
| 0 | White | **2.02** | 2.01 | 1.99 | 1.63 | 1.72 | 1.87 | 1.46 |
| | Hfchannel | **1.96** | 1.93 | 1.93 | 1.62 | 1.66 | 1.86 | 1.51 |
| | Pink | **2.10** | 2.09 | 2.08 | 1.70 | 1.68 | 1.99 | 1.58 |
| | Train | **1.68** | 1.67 | 1.61 | 1.40 | 1.31 | 1.56 | 1.43 |
| | Street | **1.64** | 1.63 | 1.67 | 1.66 | 1.38 | 1.76 | 1.51 |
| | Car | **1.77** | 1.76 | 1.85 | 1.67 | 1.46 | 1.84 | 1.55 |
| 5 | White | **2.39** | 2.38 | 2.32 | 1.85 | 2.19 | 2.16 | 1.71 |
| | Hfchannel | **2.32** | 2.30 | 2.28 | 1.93 | 2.07 | 2.15 | 1.73 |
| | Pink | **2.42** | 2.41 | 2.40 | 2.06 | 2.05 | 2.30 | 1.88 |
| | Train | **2.10** | 2.09 | 2.02 | 1.94 | 1.72 | 1.99 | 1.82 |
| | Street | **2.04** | 2.04 | 2.04 | 2.03 | 1.76 | 2.07 | 1.83 |
| | Car | **2.15** | 2.14 | 2.14 | 2.04 | 1.72 | 2.17 | 1.87 |
| 10 | White | **2.59** | 2.59 | 2.65 | 2.33 | 2.60 | 2.49 | 2.14 |
| | Hfchannel | **2.58** | 2.56 | 2.63 | 2.37 | 2.43 | 2.49 | 2.15 |
| | Pink | **2.63** | 2.63 | 2.74 | 2.48 | 2.38 | 2.57 | 2.27 |
| | Train | **2.43** | 2.41 | 2.38 | 2.33 | 2.10 | 2.36 | 2.22 |
| | Street | **2.37** | 2.35 | 2.41 | 2.41 | 2.11 | 2.41 | 2.21 |
| | Car | **2.44** | 2.43 | 2.55 | 2.45 | 2.04 | 2.50 | 2.27 |
| 15 | White | **2.72** | 2.72 | 2.961 | 2.71 | 2.97 | 2.73 | 2.50 |
| | Hfchannel | **2.72** | 2.70 | 2.93 | 2.73 | 2.78 | 2.73 | 2.49 |
| | Pink | **2.76** | 2.76 | 3.06 | 2.82 | 2.69 | 2.79 | 2.53 |
| | Train | **2.69** | 2.66 | 2.79 | 2.71 | 2.51 | 2.68 | 2.51 |
| | Street | **2.59** | 2.56 | 2.73 | 2.69 | 2.47 | 2.64 | 2.46 |
| | Car | **2.66** | 2.65 | 2.88 | 2.84 | 2.39 | 2.75 | 2.58 |

To solve the significant difference between the proposed algorithm and other traditional algorithms, we use the significance test. First, we make a hypothesis that there is no significant difference between the scores of the proposed algorithm and a traditional algorithm (including the CLSMD, MSS, SS, Subspace, MMSE, Wiener96). Then we set the value of the significance level to $\alpha = 0.05$, and the probability $p$ obtained by the significance test has two cases:

1) if $p > 0.05$, we accept the null hypothesis, i.e. there is no significant difference between the scores of the proposed algorithm and a traditional algorithm;

2) or $p < 0.05$, we reject the null hypothesis, i.e. there is a significant difference between the scores of the proposed algorithm and a traditional algorithm.

Through the data in Tables 2 and 3 of this paper, we obtain the significance test results (value of $p$) of the segSNR and the PESQ scores as shown in Tables 4 and 5. The bold numbers indicate $p < 0.05$.

By the segSNR and PESQ scores, TCCLSMD algorithm is positively significant compared to all conventional algorithms at 0 dB SNR; and TCCLSMD algorithm is more significant than MMSE and Wiener96 for all SNRs. Only by the segSNR scores and compared with the CLSMD algorithm, TCCLSMD algorithm is only positively significant at 0 dB SNR, indicating that it performs better in low SNR scenarios. Only by the PESQ scores and compared with the CLSMD algorithm, TCCLSMD algorithm is positively significant at 0 dB, 10 dB, and 15 dB SNR, indicating that it performs better in most SNR scenarios.

In summary, the ability of the proposed method to improve speech quality is more significant and stable than that of the traditional methods, especially in the case of a low SNR. It is worth emphasising that the TCCLSMD method has a better performance in improving speech quality and suppressing noise as compared with the CLSMD method.

Figure 3 shows the segSNR and PESQ score increments for the evaluation criteria values based on the TCCLSMD method minus the corresponding evaluation criteria values for the CLSMD method. It can be seen from the histogram that the most of the PESQ and segSNR incremental score values are positive.

From the aspect of PESQ measure, the proposed method has a better performance in the process of improving the speech quality of other noisy speech except for the pink noise at 15 dB SNR. From the aspect of segSNR measure, the noise suppression ability based on TCCLSMD method is greatly improved under the three noise environments of pink, train, and car, and also has a better performance under other noise environments.

Figure 4a shows the time-domain waveform comparison of enhanced speech and pure speech after different speech enhancement methods are used to process the "sp01" sampled speech signal, which is a noisy speech formed by superimposing high-frequency channel noise with 0 dB SNR. Through the visual comparison of the eight waveforms, it can be clearly seen that the spectral waveforms enhanced by the spectral subtraction (SS), MMSE, WF, and subspace method contain more burrs and have significant speech distortion. In contrast, the enhanced waveform of the TCCLSMD-based method contains few burrs, so it has a better effect in suppressing residual noise and preventing speech distortion. On the other hand, Fig. 4b shows the comparison of the speech spectrum of the pure speech, the noisy speech and the enhanced speech obtained by the TCCLSMD method. It can be clearly seen from the spectrogram that the TCCLSMD method significantly eliminates more noise information and retains more speech information. The spectral map of this proposed method is very close to the "clean speech" signal spectrogram.

Table 4. Significance test between the TCCLSMD and a traditional algorithm for the seg-SNR scores.

| SNR [dB] | CLSMD | MSS | SS | Subspace | MMSE | Wiener96 |
|---|---|---|---|---|---|---|
| 0 | **0.0387** | **0.0384** | **0.0408** | **0.0008** | **0.0092** | **0.0425** |
| 5 | 0.1260 | 0.1469 | 0.0518 | **0.0020** | **0.0010** | **0.0145** |
| 10 | 0.1573 | **0.0240** | 0.3770 | 0.1499 | **0.0007** | **0.0051** |
| 15 | 0.1293 | **0.0001** | **0.0418** | **0.0042** | **0.0003** | **0.0015** |

Table 5. Significance test between the TCCLSMD and a traditional algorithm for the PESQ scores.

| PESQ [dB] | CLSMD | MSS | SS | Subspace | MMSE | Wiener96 |
|---|---|---|---|---|---|---|
| 0 | **0.0113** | **0.0004** | **9.8740e–07** | **0.0007** | **3.8174e–05** | **1.4480e–06** |
| 5 | 0.5178 | 0.1249 | **7.4463e–05** | **0.0226** | **8.0606e–05** | **1.5736e–05** |
| 10 | **0.0252** | **0.0010** | **4.0430e–05** | 0.2421 | **0.0010** | **0.0049** |
| 15 | **0.0014** | **0.0001** | **7.1446e–07** | **0.0008** | **8.0606e–05** | **0.0025** |

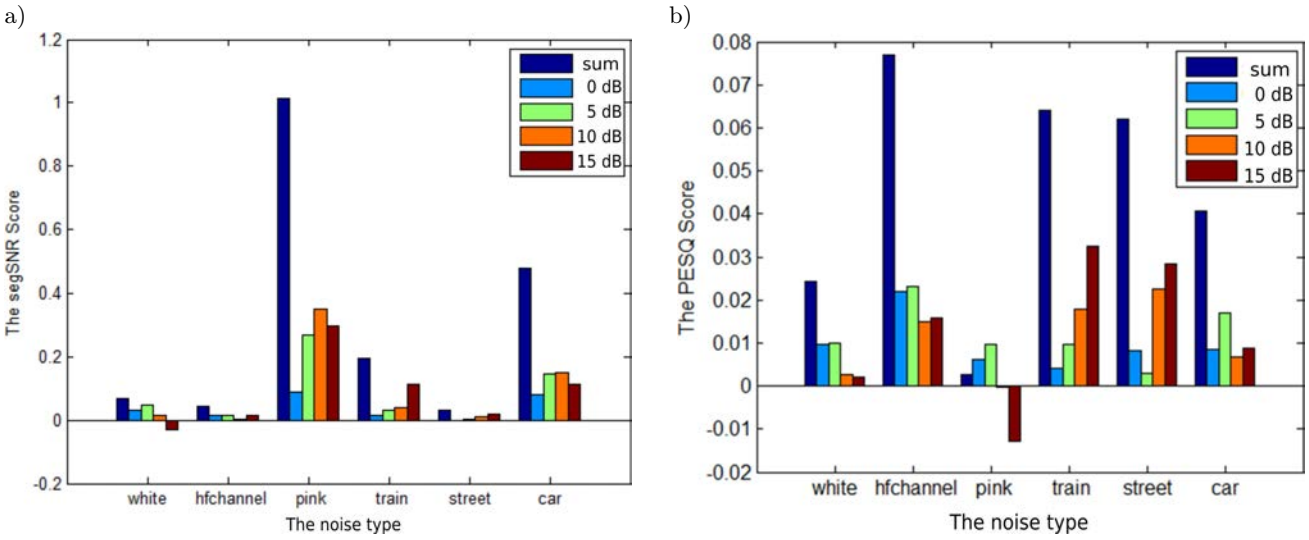


Fig. 3. TCCLSMD vs. CLSMD for segSNR and PESQ incremental scores: a) the scores for segSNR, b) the scores for PESQ.

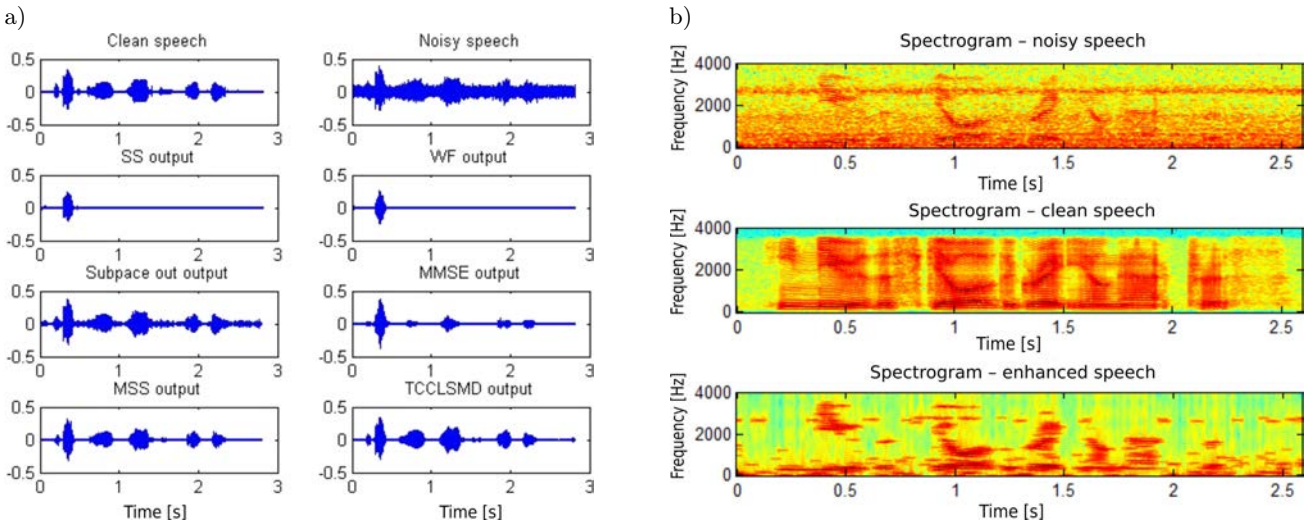


Fig. 4. Waveform and Spectral comparisons of speech corrupted by noise at 0 dB SNR: a) waveform comparison of speech ("sp01") corrupted by hfchannel, b) spectral comparison of speech ("sp02") corrupted by f16.

## 5. Conclusions

In this paper, we propose a temporal continuity constrained low-rank sparse matrix decomposition (TCCLSMD) based speech enhancement method. First, we introduce temporal continuity restriction and sparse matrix reconstruction model, and describe the algorithms based on the TCCLSMD method. Then, we build up the TCCLSMD speech enhancement system based on the AMS framework. Finally, we compare the proposed method with the traditional speech enhancement methods, and find that the ability of the proposed method to improve speech quality is more significant and stable than that of the traditional methods, especially in the case of low SNR. Moreover, the presented method can directly obtain noise and speech signal information through matrix decomposition, and does not need pre-voice detection processing. It can make good use of the long-term statistical characteristics of background noise signals, and it is easy to obtain stable noise reduction effect. In summary, increasing the temporal continuity constraint can better achieve the purpose of speech enhancement.

## Acknowledgments

This work was supported by National Natural Science Foundation of China (No. 61861033, 61761031, 61866027, 61876213, 61401259), Jiangxi Province Natural Science Foundation (No. 20181BAB202022), Jiangxi Provincial Department of Education Foundation (No. GJJ170599, JXJG-17-8-12), and Research Fund for Young and Middle-aged Scientists of Shandong Province (No. ZR2016FB25).

## References


1. Abdali S., NaserSharif B. (2017), *Non-negative matrix factorization for speech/music separation using source dependent decomposition rank, temporal continuity term and filtering*, Biomedical Signal Processing and Control, **36**, 168–175, doi: 10.1016/j.bspc.2017.03.010.
2. Bando Y. *et al.* (2018), *Speech enhancement based on Bayesian low-rank and sparse decomposition of multichannel magnitude spectrograms*, IEEE/ACM Transactions on Audio, Speech, and Language Processing, **26**, 2, 215–230, doi: 10.1109/TASLP.2017.2772340.
3. Boll S.F. (1979), *Suppression of acoustic noise in speech using spectral subtraction*, IEEE Transactions on Audio, Speech, and Signal Processing, **27**, 2, 113–120, doi: 10.1109/TASSP.1979.1163209.
4. Bouwmans T., Sobral A., Javed S., Jung S.K., Zahzah E.-H. (2017), *Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset*, Computer Science Review, **23**, 1–71, doi: 10.1016/j.cosrev.2016.11.001.
5. Cai J.F., Candčs E.J., Shen Z. (2010), *A singular value thresholding algorithm for matrix completion*, SIAM Journal on Optimization, **20**, 4, 1956–1982, doi: 10.1137/080738970.
6. Candes E.J., Li X., Ma Y., Wright J. (2011), *Robust principal component analysis?*, Journal of the ACM, **58**, 3, 1–37, doi: 10.1145/1970392.1970395.
7. Candes E.J., Plan Y. (2010), *Matrix completion with noise*, Proceedings of the IEEE, **98**, 6, 925–936, doi: 10.1109/JPROC.2009.2035722.
8. Cohen I. (2004), *Speech enhancement using a noncausal a priori SNR estimator*, IEEE Signal Processing Letters, **11**, 9, 725–728, doi: 10.1109/LSP.2004.833478.
9. Ephraim Y., Van Trees H. (1995), *A signal subspace approach for speech enhancement*, IEEE Transactions on Speech and Audio Processing, **3**, 4, 251–266, doi: 10.1109/89.397090.
10. Hermus K., Wambacq P., Hamme H.V. (2007), *A review of signal subspace speech enhancement and its application to noise robust speech recognition*, EURASIP Journal on Advances in Signal Processing, **2007**, 1, 045821, doi: 10.1155/2007/45821.
11. Hu Y., Loizou P.C. (2003), *A generalized subspace approach for enhancing speech corrupted by colored noise*, IEEE Transactions on Audio, Speech and Language Processing, **11**, 4, 334–342, doi: 10.1109/TSA.2003.814458.
12. Hu Y., Loizou P.C. (2008), *Evaluation of objective quality measures for speech enhancement*, IEEE Transactions on Audio, Speech and Language Processing, **16**, 1, 229–230, doi: 10.1109/TASL.2007.911054.
13. Jin K.H., Ye J.C. (2018),*Sparse and low-rank decomposition of a hankel structured matrix for impulse noise removal*, IEEE Transactions on Image Processing, **27**, 3, 1448–1461, doi: 10.1109/TIP.2017.2771471.
14. Kammi S., Mollaei M.R.K. (2017), *Noisy speech enhancement with sparsity regularization*, Speech Communication, **87**, 58–69, doi: 10.1016/j.specom.2017.01.003.
15. Kheder W.B., Matrouf D., Bousquet P.-M., Bonastre J.-F., Ajili M. (2017), *Fast i-vector denoising using MAP estimation and a noise distributions database for robust speaker recognition*, Computer Speech & Language, **45**, 104–122, doi: 10.1016/j.csl.2016.12.007.
16. Kolbæk M., Tan Z.-H., Jensen J. (2017), *Speech intelligibility potential of general and specialized deep neural network based speech enhancement systems*, IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), **25**, 1, 153–167, doi: 10.1109/TASLP.2016.2628641.
17. Li X., Fan M., Liu L., Li W. (2018), *Distributed-microphones based in-vehicle speech enhancement via sparse and low-rank spectrogram decomposition*, Speech Communication, **98**, 51–62, doi: 10.1016/j.specom.2017.12.008.
18. Liu H., Peng J. (2018), *Sparse signal recovery via alternating projection method*, Signal Processing, **143**, 161–170, doi: 10.1016/j.sigpro.2017.09.003.
19. Loizou P.C. (2007), *Speech Enhancement: Theory and Practice*, New York: Taylor & Francis.
20. Lu Y., Loizou P.C. (2008), *A geometric approach to spectral subtraction*, Speech Communication, **50**, 6, 453–466, doi: 10.1016/j.specom.2008.01.003.
21. Mavaddaty S., Ahadi S. M., Seyedin S. (2016), *A novel speech enhancement method by learnable sparse and low-rank decompositionand domain adaptation*, Speech Communication, **76**, 42–60, doi: 10.1016/j.specom.2015.11.003.
22. Mohammadiha N., Arne L. (2013), *Nonnegative HMM for babble noise derived from speech HMM: Application to speech enhancement*, IEEE Transactions on Audio, Speech, and Language Processing, **21**, 5, 998–1011, doi: 10.1109/TASL.2013.2243435.
23. Moor, de B. (1993), *The singular value decomposition and long and short spaces of noisy matrices*, IEEE Transactions on Signal Processing, **41**, 9, 2826–2839, doi: 10.1109/78.236505.
24. Paliwal K., Schwerin B., Wójcicki K. (2012), *Speech enhancement using a minimum mean-square error short-time spectral modulation magnitude estimator*, Speech Communication, **54**, 2, 282–305, doi: 10.1016/j.specom.2011.09.003.
25. Paliwal K., Wójcicki K., Schwerin B. (2010), *Single-channel speech enhancement using spectral subtraction in the short-time modulation domain*, Speech Communication, **52**, 5, 450–475, doi: 10.1016/j.specom.2010.02.004.
26. Plapous C., Marro C., Scalart P. (2006), *Improved signal-to-noise ratio estimation for speech enhancement*, IEEE Transactions on Acoustics, Speech, and Signal Processing, **14**, 6, 2098–2108, doi: 10.1109/TASL.2006.872621.

27. Quatieri T. (2002), *Discrete-time speech signal processing: principles and practice*, Prentice Hall, Upper Saddle River, NJ.

28. Rugini L., Banelli P. (2016), *On the equivalence of maximum SNR and MMSE estimation: applications to additive non-Gaussian channels and quantized observations*, IEEE Transactions on Signal Processing, **64**, 23, 6190–6199, doi: 10.1109/TSP.2016.2607152.

29. Scalart P., Vieira-Filho J. (1996), *Speech enhancement based on a priori signal to noise estimation.* Proceedings on 21st IEEE International Conference on Acoustics, Speech, and Signal Processing Conference, Atlanta, GA, doi: 10.1109/ICASSP.1996.543199.

30. Shannon B., Paliwal K. (2006), *Role of phase estimation in speech enhancement*, [in:] *INTERSPEECH-2006*, paper 1330-Tue3FoP.4, https://www.isca-speech.org/archive/archive_papers/interspeech_2006/i06_1330.pdf.

31. Shi J., Song W. (2016), *Sparse principal component analysis with measurement errors*, Journal of Statistical Planning and Inference, **175**, 87–99, doi: 10.1016/j.jspi.2016.03.001.

32. Stark A., Paliwal K. (2011), *Use of speech presence uncertainty with MMSE spectral energy estimation for robust automatic speech recognition*, Speech Communication, **53**, 1, 51–61, 10.1016/j.specom.2010.08.001.

33. Sun C., Mu J. (2015), *An eigenvalue filtering based subspace approach for speech enhancement*, Noise Control Engineering Journal, **63**, 1, 36–48, doi: 10.3397/1/376305.

34. Sun C., Xie J., Leng Y. (2016), *A signal subspace speech enhancement approach based on joint low-rank and sparse matrix decomposition*, Archives of Acoustics, **41**, 2, 245–254, 10.1515/aoa-2016-0024.

35. Sun C., Zhu Q., Wan M. (2014), *A novel speech enhancement method based on constrained low-rank and sparse matrix decomposition*, Speech Communication, **60**, 44–55, doi: 10.1016/j.specom.2014.03.002.

36. Sun M., Li Y., Gemmeke J.F., Zhang X. (2015), *Speech enhancement under low SNR conditions via noise estimation using sparse and low-rank NMF with Kullback-Leibler divergence*, IEEE Transactions on Audio, Speech, and Language Processing, **23**, 7, 1233–1242, doi: 10.1109/TASLP.2015.2427520.

37. Tan H., Cheng B., Feng J., Feng G., Wang W., Zhang Y.-J. (2013), *Low-n-rank tensor recovery based on multi-linear augmented Lagrange multiplier method*, Neurocomputing, **119**, 144–152, doi: 10.1016/j.neucom.2012.03.039.

38. Virtanen T. (2007), *Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria*, IEEE Transactions on Audio, Speech, and Language Processing, **15**, 3, 1066–1074, doi: 10.1109/TASL.2006.885253.

39. Wiener N. (1949), *Extrapolation, interpolation, and smoothing of stationary time series*, New York: Wiley.

40. Wright J., , Peng Y., Ma Y., Ganesh A., Rao S. (2009), *Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization*, [in:] *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J.D. Lafferty, C.K.I. Williams, A. Culotta [Eds], pp. 2080–2088, http://papers.nips.cc/paper/3704-robust-principal-component-analysis-exact-recovery-of-corrupted-low-rank-matrices-via-convex-optimization.pdf.

41. Xu H., Caramanis C., Sanghavi S. (2012), *Robust PCA via outlier pursuit*, IEEE Transactions on Information Theory, **58**, 5, 3047–3064, doi: 10.1109/TIT.2011.2173156.

42. Zhang Y., Zhao Y. (2013), *Real and imaginary modulation spectral subtraction for speech enhancement*, Speech Communication, **55**, 4, 509–522, doi: 10.1016/j.specom.2012.09.005.

43. Zhen L., Peng D., Yi Z., Xiang Y., Chen P. (2017), *Underdetermined blind source separation using sparse coding*, IEEE Transactions on Neural Networks and Learning Systems, **28**, 12, 3102–3108, 10.1109/TNNLS.2016.2610960.