DOI: 10.1051/matecconf/201712801003

Speech Denoising in White Noise Based on Signal Subspace Low-rank Plus Sparse Decomposition

Shuai yuan[1], Cheng-li Sun[2,a]

[1]*Science and Technology on Avionics Integration Laboratory, Shanghai, China*
[2]*Nanchang Hangkong University, Nanchang, 330063 China*

Abstract—In this paper, a new subspace speech enhancement method using low-rank and sparse decomposition is presented. In the proposed method, we firstly structure the corrupted data as a Toeplitz matrix and estimate its effective rank for the underlying human speech signal. Then the low-rank and sparse decomposition is performed with the guidance of speech rank value to remove the noise. Extensive experiments have been carried out in white Gaussian noise condition, and experimental results show the proposed method performs better than conventional speech enhancement methods, in terms of yielding less residual noise and lower speech distortion.

Keywords- speech enhancement; subspace method; low-rank plus sparse decomposition.

1. Introduction

Speech enhancement refers to the improvement in quality and intelligibility of noise corrupted speech signals by using supervised or unsupervised speech enhancement methods. It is widely used as a pre-processing block in a lot of applications like automatic speech recognizer and other communication systems.

Over the last fifty decades, many algorithms have been proposed about for speech enhancement. The typical algorithms including spectral subtraction[1], minimum mean square error (MMSE) estimation [2-4], Wiener filtering [5-8], and subspace methods [9-13]. Spectral Subtraction and Wiener filtering have been widely used for enhancing speech because of their simplicity and ease of implementation in single channel systems but they suffer from the production of musical noise after enhancement and is one of their major drawbacks. Signal subspace approach [9-13], have shown to give a better compromise between less residual noise and signal distortion of the output signal, compared to the other existing techniques.

Signal subspace approach was firstly proposed by Ephraim Y, et al. The principle of this method is to separate the noisy speech observation space into a signal subspace and a noise subspace, and the enhanced speech was constructed using only the components of the signal within the signal subspace. In the subspace-based algorithms, subspace decomposition is a critical step for subspace separation, which is often performed via Karhunen-Loeve transform (KLT)[10] or singular value decomposition (SVD) [9]. The main issue in developing a subspace-based model is the way of splitting and refining the signal and noise subspace in an optimal way. In [14], variance of the reconstruction error criterion was introduced to optimize the subspace selection for speech enhancement. In [15], to optimize the subspace decomposition model, human auditory psychoacoustic properties are incorporated into the subspace filter to reconstruct the enhanced signal. Although many efforts were conducts to improve the subspace methods, the existing subspace-based speech enhancement methods still suffer from the problem of low decomposition accuracy in the presence of large noise, resulting in a high remainder noise within enhanced speech in strong noise cases.

In this paper, we propose a new subspace-based method for speech enhancement based on the principle of low-rank and sparse decomposition (LSD). The main idea behind our method is motivated by the recent development of low-rank and sparse theory [16]. According to this theory, if a given corrupted data matrix Y has an underlying low-rank structure, yet corrupted by sparse additive noises. The underlying low-rank component L can be effectively recovered by solving a convex optimization problem, even if the noise is arbitrary in magnitude. In the time domain, owing to the short-time stability of human speech, speech signals can be assumed to have a low-rank structure. On the other hand, due to the randomness of noise, background noise is more variable and thus can be viewed as sparse and high-rank. Thus LSD theory can be exploited to recover the underlying speech from corrupted speech signals.

The rest of the paper is organized as follows. We first briefly review the previous works in Section 2. In Section 3, we describe the LSD based signal subspace speech enhancement method. Section 4 presents the experiments and results. Finally, we give the concludes and future work in section 5.

2. Related work

The goal of principal component analysis (PCA) technique is to determine the most significant basis to re-express a

[a] Corresponding author: sun_chengli@163.com

noisy speech set [17]. This new basis will filter out the noise and reduce a multidimensional speech to lower dimensions by avoiding redundant data.

Let us consider the problem of the enhancement of a speech signal contaminated by an independent additive noise. Let $x(t)$ and $d(t)$ denote the sampled clean speech and noise signal, respectively. The observed noisy speech signal $y(t)$ is

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

Suppose $y(t)$ was framed with the length N. Arranging the N-dimensional vectors into a $(M\text{-l}+1)\times l$ Toeplitz structure matrix, we can get

$$Y = X + D. \tag{2}$$

Assuming that the rank of matrix Y is r, the optimal enhanced speech matrix $\hat{X}$ can be estimated according to the following least-square criterion

$$\min_{\hat{X}} \left\| Y - \hat{X} \right\|_F^2, \ \ \operatorname{rank}(\hat{X}) \le r \ \ , \tag{3}$$

where symbol $\|\cdot\|_F$ denotes the Frobenius norm of a matrix and $\|X\|_F = \sqrt{X_{ij}^2}$.

If $d(t)$ is a white Gaussian noise, it satisfies the conditions $D^T D = \sigma_d^2 I$ and $X^T D = 0$. Where σ_d^2 is the variance of noise. The optimal solution of (4) can be obtained by applying singular value decomposition (SVD) of Y.

$$Y = U\Sigma V^T$$

$$\hat{X} = \sum_{i=1}^{r} \lambda_i U_i V_i^T . \tag{4}$$

Here, U and V are two orthogonal matrices holding the left and right (approximate) singular vectors of given matrix, and Λ is a diagonal matrix holding the singular values: $\lambda_1 \ge \lambda_2 \ge \cdots \lambda_{r-1} \ge \lambda_r$.

The above low-rank matrix $\hat{X}$ represents the original speech matrix X in the sense of least-square minimization. This may get the optimal estimate when the noise is small, independent, and identically distributed Gaussian.

However, PCA is highly sensitive to the presence of large corruptions. Even a single outlier in the data matrix can render the estimation of the low-rank component arbitrarily far from the true model. In [16], a new theory called Robust PCA was developed for this shortcoming. The basic idea of Robust PCA is to decompose the data matrix M as M=L+S, where S Î ¡ $^{N' K}$ is a sparse matrix with a sparse number of non-zero coefficients with arbitrarily large magnitude. RPCA can be solved by minimizing the following convex program

$$\min \|L\|_* + \lambda \|S\|_1 , \text{ s.t. } M = L + S, \tag{5}$$

where $\|\cdot\|_*$ denotes the matrix nuclear norm, which is defined as the sum of all singular values and is suggested as a convex surrogate to the rank function [18]. $\|\cdot\|_1$ denotes the l_1-norm of a matrix, which is defined as the sum of the absolute values of matrix elements. This problem is known to have a stable solution provided L and S are sufficiently incoherent [19], i. e., the low-rank matrix is not sparse and the sparse matrix is not low-rank. More recently, RPCA theory was introduced into the speech enhancement task in [20], where a constrained low-rank and sparse matrix decomposition (CLSMD) algorithm is designed for noise reduction.

3. LSD based speech denoising method

In this work, we propose a new subspace decomposition algorithm based on the LSD, which is less sensitive to the large noise interferences.

Firstly, we formulate the speech enhancement problem as the following optimization problem,

$$\begin{aligned} &\min_{\mathbf{L,S}} \|Y - L - S\|_F^2 , \\ &s.t. \ \ rank(L) \le r, \ \ |S|_0 \le h. \end{aligned} \tag{6}$$

The above formula can be solved by alternatively solving the following two formulas until convergence

$$\begin{cases} L_i = \underset{\operatorname{rank}(L)\le r}{\arg\min} \|Y - L - S_{i-1}\|_F^2 & \text{(a)} \\ S_i = \underset{|S|_0 \le h}{\arg\min} \|Y - L_i - S\|_F^2 & \text{(b)} \end{cases} \tag{7}$$

Given an estimate of sparse matrix S_{i-1}, the minimization in (7-a) over L is to learn a rank-r low-rank matrix from partial observations. This is a fixed-rank approximation problem, we can solve it use bilateral random projections (BRP) based fast low-rank matrix approximation.

$$L_t = M_1 (A_2^T M_1)^{-1} M_2^T \tag{8}$$

Where $M_1 = YA_1$, $M_2 = Y^T A_2$. Both $A_1 \in R^{n\times r}$ and $A_2 \in R^{m\times r}$ are Gaussian random matrices.

The minimization in (7-b) over S is to learn a sparse matrix from partial observations. This can be computed via entry-wise hard thresholding function [21],

$$\varphi_T(x) = x \cdot 1(|x| > u), \tag{9}$$

which keeps the input if it is larger than the threshold; otherwise, it is set to zero. In summary, we have following optimization algorithm for LSD.

Algorithm 1. Optimization algorithm for LSD

```
 Given $r$, $T$, $\varepsilon$, $t_{\text{maxiter}}$;
 Initialize $Y_0 = Y$ , $S_t = [0]_{N\times K}$ , $t$=0;
while not converged do
      %Update of low-rank matrix  $L$
      $A_1 = randn(n,r)$ ;
      $A_2 = randn(m,r)$
      $M_1 = Y_t A_1$,
      $M_2 = Y_t^T A_2$
      $L_t = M_1 (A_2^T M_1)^{-1} M_2^T$ ;
      %Update of   sparse matrix $S$
      $X_t = Y_t - L_t + S_t$ ;
       $S_t = X_t \otimes (X_t > T)$ ;
       % Stopping criteria
       If  $\|Y - L_t - S_t\|_F^2 / \|Y\|_F^2 \le \varepsilon$ or $t$= = $t_{\text{maxiter}}$
          break;
       end
       $Y_t = L_t + X_t - S_t$
       $t = t + 1$;
```

```
end while
output: $L= L_t$ , $S= S_t$
```

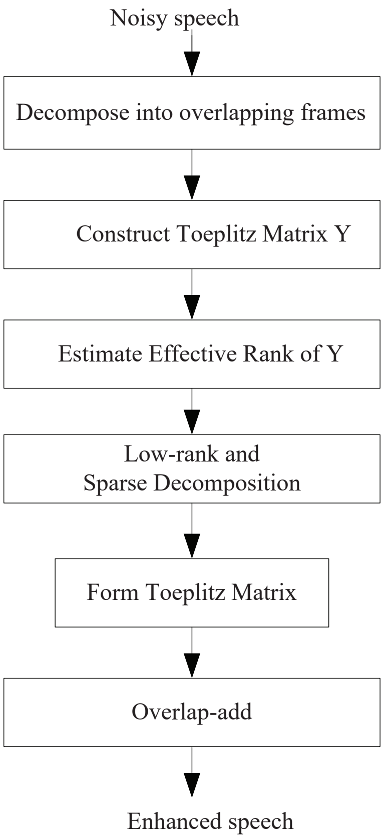

Figure 1. The scheme of LSD based speech enhancement method

Figure 1 shows the scheme of LSD based speech enhancement method. At first, the noisy speech signal is divided into frames in the time domain. Then we arrange each frame of the noisy speech into a Toeplitz matrix. After we estimated the effective rank r with the analysis-by-synthesis approach [22], the noisy speech matrix Y is decomposed into the low-rank matrix L with the rank r using the LSD algorithm. Since L is not a Toeplitz matrix, we average all the diagonal elements of L to let it became a Toeplitz matrix form. Finally, the enhanced speech is constructed by taking the inverse transform of Toeplitz matrix followed by least-squares overlap-add synthesis [23].

4. Experimental results

For evaluation of the proposed JLSMD method, we choose a total of 30 sentences (sp01~sp30) taken from NOIZEUS database. Both speech and noise were sampled at 8 kHz 16 bits. Time frame length is 264 sample points with 50% frame overlap. White Gaussian noise was added to clean speech at various levels. We use segSNR and PESQ ((Perceptual Evaluation of Speech Quality) scores for performance measure. four conventional speech enhancement methods: spectral subtraction (SSboll [1]), Subspace SVD based subspace decomposition algorithm (SSVD) [9], Wiener filter based method (Wiener [8]), minimum mean-square error algorithm (MMSE [24]), KLT [12] and CLSMD [20]).

Tables 1 and 2 show the comparison of performance in terms of PESQ and segSNR scores. The larger the PESQ-MOS and segSNR scores are, the better the performances are. We can see the proposed method LSD has got the highest PESQ-MOS and segSNR scores among all the compared methods, except at 0 dB where CLSMD has the highest segSNR score.

Table 1. PESQ scores in the white noise case at different SNRs

Methods	0 dB	5 dB	10 dB	15 dB
KLT	1.100	3.529	5.937	8.058
MMSS	-0.464	0.787	2.119	3.374
SSboll	-3.519	-2.213	-1.061	0.051
SSVD	0.4106	2.9652	5.4880	7.9579
CLSMD	2.146	3.685	4.856	5.640
LSD	1.468	3.970	6.596	9.074

Table 2. PESQ scores in the white noise case at different SNRs

Methods	0 dB	5 dB	10 dB	15 dB
KLT	1.994	2.397	2.744	3.078
MMSS	1.455	1.710	2.140	2.498
SSboll	1.651	1.943	2.179	2.462
SSVD	1.713	2.194	2.596	2.972
CLSMD	2.009	2.384	2.587	2.720
LSD	2.055	2.478	2.844	3.198

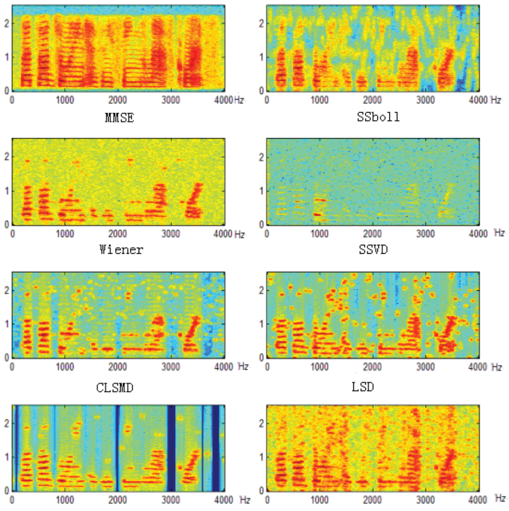

Figure 2. Comparison of the spectrograms for speech enhanced by different methods

Fig. 2 presents spectrogram comparisons for various speech enhancement methods in the 10 dB SNR. We can see from these enhanced speech spectrograms. Along with the high levels of noise reduction, the proposed LSD based method is still able to preserve most of the low-energy speech components compared with the seven speech enhancement methods.

5. Conclusions

In this paper, we presented a LSD based signal subspace speech enhancement method. The proposed method is less sensitive to the large interferences as compared with traditional algorithms, and can significantly reduce noise. Experiments demonstrate that the proposed method is good at improving the overall enhanced speech quality, especially in low SNRs. It should be pointed out that LSD method has improved the original subspace method based on SVD and can wipe out more residual noise. In the future research work we will devote more efforts to improving the noise reduction performancein the colored noise.

Acknowledgements

This work was supported by China's Aviation Science fund (No. 20145556011).

References

1. Boll, S.F., Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust. Speech Signal Process, 1979. 27(2): p. 113-120.
2. Ephraim, Y. and D. Malah, Speech enhancement using a minimum mean square error log-spectral amplitude estimator. IEEE Trans. Acoust., Speech, Signal Process, 1985. SSP-33,(2): p. 443–445.
3. Ephraim, Y. and D. Malah, Speech enhancement using a minimum mean square error short-time spectral amplitude estimator. IEEE Transactions on Acoustics, Speech, and SignalProcessing, 1984. vol. ASSP-32(6): p. 1109-1121.
4. Stark, A. and K. Paliwal, Use of speech presence uncertainty with MMSE spectral energy estimation for robust automatic speech recognition. Speech Communication, 2011(53): p. 51-61.
5. Soon, I.Y. and S.N. Koh. Low distortion speech enhancement. in Inst. Elec. Eng. 2000.
6. Wiener, N., Extrapolation, Interpolation, and Smoothing of Stationary Time Series. 1949, New York: Wiley.
7. Plapous, C., C. Marro, and P. Scalart, Improved Signal-to-Noise Ratio Estimation for Speech Enhancement IEEE Transactions on Acoustics, Speech, and SignalProcessing, 2006. 14(6): p. 2098-2108.
8. Scalart, P. and J. Vieira-Filho, Speech enhancement based on a priori signal to noise estimation, in Proc. 21st IEEE Int. Conf. Acoust. SpeechSignal Processing. 1996: Atlanta, GA. p. 629-632.
9. Moor, B.D., The singular value decomposition and long and short spaces of noisy matrices. IEEE Trans. on Signal Processing, 1993. 41(9): p. 2826-2839.
10. Ephraim, Y. and H. Van Trees, A signal subspace approach for speech enhancement. IEEE Trans. Speech Audio Process., 1995. 3(4): p. 251-266.
11. Doclo, S. and M. Moonen, GSVD-Based Optimal Filtering for Single and Multimicrophone Speech Enhancement. IEEE Transactions on signal processing, 2002. 50(9): p. 2230-2242.
12. Hu, Y. and P.C. Loizou, A Generalized Subspace Approach for Enhancing Speech Corrupted by Colored Noise. IEEE Transactions on Audio, Speech and Language Processing 2003. 11(4): p. 334-342.
13. Hermus, K., P. Wambacq, and H.V. hamme, A Review of Signal Subspace Speech Enhancement and Its Application to Noise Robust Speech Recognition. EURASIP Journal on Advances in Signal Processing, 2007: p. 1-15.
14. Saadoune, A., A. Amrouche, and S.-A. Selouani, Perceptual subspace speech enhancement using variance of the reconstruction error. Digital Signal Processing, 2014. 22: p. 187-196.
15. Surendran, S. and T.K. Kumar, Variance normalized perceptual subspace speech enhancement. International Journal of Electronics and Communications, 2017. 74: p. 44-54.
16. Wright, J., Y. Peng, and Y. Ma, Robust Principal Component Analysis: Exact Recovery of Corrupted Low-rank Matrices by Convex Optimization. In NIPS, 2009.
17. Jolliffe, I.T., Principal Component Analysis. Springer Series in Statistics. 2002, New York: Springer.
18. Candes, E.J. and T. Terence, The power of convex relaxation: near-optimal matrix completion. IEEE Transactions on Information Theory, 2010. 56(5): p. 2053-2080.
19. Candes, E.J., et al., Robust Principal Component Analysis? Journal of the ACM, 2011. 58(3): p. 1-37.
20. Sun, C., Q. Zhu, and M. Wan, A novel speech enhancement method based on constrained low-rank and sparse matrix decomposition. Speech Communication, 2013. 60(12): p. 44-55.
21. Chang, S.G., B. Yu, and M. Vetterli, Adaptive Wavelet Thresholding for Image Denoising and Compression. IEEE Transactions on Information Theory, 2000. 9(9): p. 1532-1547.
22. Loizou, P.C., Speech Enhancement: Theory and Practice. 2007, New York: Taylor & Francis.
23. Quatieri, T., Discrete-Time Speech Signal Processing: Principles and Practice. 2002, Prentice Hall, Upper Saddle River, NJ.
24. Cohen, I., Speech Enhancement Using a Noncausal A Priori SNR Estimator. IEEE Signal Processing Letters, 2004. 11(9): p. 725-728.