A New Sparse Representation Algorithm for Speech Denoising

Yan Zhou
School of Electronic Information Engineering
Suzhou Vocational University
Suzhou Jiangsu, China
zhyan@jssvc.edu.cn

***Abstract*— This paper proposes a new speech denoising method that uses K-SVD sparse representation algorithm. This approach is based on sparse and redundant representation over dictionary. Here, spectrogram patches are used as training samples for the initial redundant dictionary. However, since the K-SVD algorithm is limited in handling small size spectrogram, the training samples need to extend their deployment to arbitrary spectrogram sizes by defining a global spectrogram prior that forces sparsity over patches in every location in the spectrogram. Simulation experiments show that the performance of the proposed K-SVD denoising algorithm is stable, and the white noise can be effectively separated. In addition, K-SVD algorithm is a simple and effective algorithm which surpasses the redundant DCT method and Gabor dictionary. In a word, K-SVD algorithm leads to an alternative speech denoising method.**

Keywords- speech denoising, spectrogram, K-SVD algorithm, redundant dictionary, sparse representation

I. INTRODUCTION

Speech is one of the simplest ways for signal expression. The research of speech denoising is important, because of the evident applications it serves. At present, eliminating the background noise and improving the speech quality become a significant research direction. Recently, the speech denoising method such as the traditional spectral subtraction method [1-2], wavelet transform method [3-4], linear forecast analysis [5], the subspace denoising method [6], and various improvement denoising algorithm[7-8]. However, these methods exit a lot of defects. In recent years, with the growing realization of sparse representation theory, the speech denoising method based on signal sparse representation has drawn a lot of research attention [9-10]. Sparse decomposition method starts from the point of non-stationary characteristic of speech signal. It realizes the speech denoising by analyzing the time-frequency characteristic of speech signal reasonably. Because the K-SVD algorithm [11-12] is simple, flexible and efficient, moreover, K-SVD redundant dictionary is produced by training speech samples, moreover, it can reflect the inherent structure characteristics of speech signal itself. Thus, a speech denoising algorithm that is via sparse and redundant representation over K-SVD dictionary is proposed in this paper.

II. PROPOSED K-SVD ALGORITHM

The speech redundant dictionary training based on K-SVD learning algorithm includes two steps, one is sparse decomposition, and the other is dictionary atoms updating. When considering this available scheme, the two steps should be alternately performed, thus the redundant dictionary and the sparse matrix can be updated synchronously. Generally speaking, the target of K-SVD learning algorithm is to realize sparse representation adaptively. That is, the signal similar with the training signal can be sparse represented over this redundant dictionary. The specific steps of speech denoising algorithm based on K-SVD redundant dictionary are described as following.

Step1 Initialization: The redundant DCT dictionary is employed as the initial dictionary $D=\left\{d_j\right\}_{j=1}^{K}$, and dictionary redundancy is set for γ . The training samples and testing samples all should be transformed to corresponding spectrogram. The size of each spectrogram is set as $p\times p$, and doing pretreatment for this spectrograms before dictionary training. Since dictionary learning is limited in handling small spectrogram patches, in this work, a global spectrogram prior that forces sparsity over patches in every location in the spectrogram (with overlaps) is proposed. So the spectrograms ought to be divided into small overlapping patches randomly.

Step2 Sparse Coding: In this step, the initial dictionary must be assumed as fixed firstly. Aligning with this, the corrupted speech signal $Y=\left\{y_i\right\}_{i=1}^{N}$ should be decomposed over this initial dictionary by using orthogonal matching algorithm (OMP). Since the iterative termination condition is decided by the decomposition residual .Consequently, the sparse matrix can be got by solving the following formula.

$$\min_{D.X}\sum_i\left\|x_i\right\|_0 \qquad s.t\left\|Y-DX\right\|_F^2\le\varepsilon \tag{1}$$

Step3 Dictionary updating: This step is aim to update the redundant dictionary. It is important to fix the sparse matrix X which has been trained at the last step. For the size of K-SVD dictionary is $n\times k$, it is updated column by column and this is an iterative process. Meanwhile, the error E_k is calculated in all samples which have been removed

the composition of atom d_k . Notice that error E_k can be calculated as:

$$E_k = Y - \sum_{k=j} d_k x_T^k \qquad (2)$$

Here, x_T^k is the k-th line in sparse matrix X of d_k . So that the column error E_R^k of atom d_k can be expressed as:

$$E_R^k = E_k \Omega_k \qquad (3)$$

Here, Ω_k is the matrix $N \times |w_k|$, $w_k = \{ i | 1 \le i \le N, x_T^k \ne 0 \}$, w_k is the index group for all the samples been decomposed which use atom d_k . Then the E_R^k needs to do SVD decomposition, the calculation formula is defined as

$$E_R^k = U \Delta V^T \qquad (4)$$

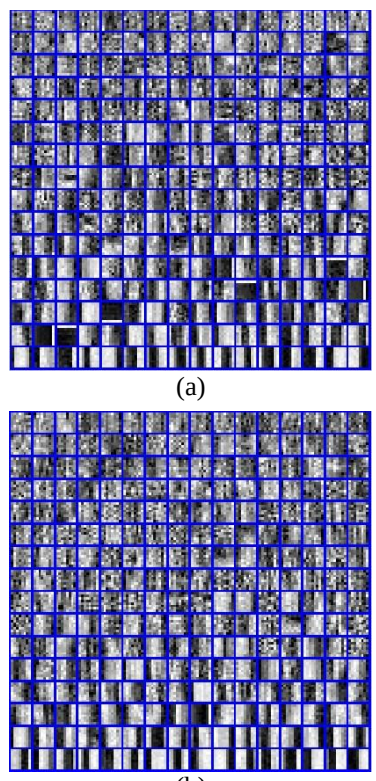
Thus, the first column of U is the updated atom $\tilde{d}_k$. Taken these steps, the updated dictionary can be obtained until it reaches iterative termination condition. When the iterative stops, the updated dictionary is fixed again, it is ready for doing sparse decomposition in the next step. The adaptive redundant K-SVD dictionary can be gotten by repeating step2 and step3.After acquiring the redundant dictionary, OMP algorithm is employed to do sparse decomposition for gaining the sparse coefficients. By this way, the useful spectrogram signal can be reconstructed using these sparse coefficients but except the noise spectrograms, thus, the useful speech part and the noise part can be divided, in short, the whole spectrogram quality can be enhanced.

III. EXPERIMENTAL RESULTS AND ANALYSIS

In order to validate the validity and advantage of the K-SVD algorithm in the progress of speech denoising, the experiments are designed as following.

A. Anti-noise performance of the K-SVD dictionary

When the denoising task is adopted at hand, a crucial step is the choice of the examples to train on. The choosing of training sample is also one of the important steps in the denoising algorithm based on redundant K-SVD dictionary. The experiments following conclude that while a reasonably good dictionary that fits all is indeed within reach. Furthermore, the appropriate sample is related to the performance of the dictionary. In order to find the influence factors of training samples on the K-SVD algorithm, the experiments are performed use different training samples. In all this experiments, redundant DCT dictionary is used for the initial training dictionary. The OMP algorithm is introduced for sparse decomposition. Moreover, the dictionary redundancy is set as 4. The experiment starts with training clean spectrograms. Choosing 5 train samples and the duration for each sample is 0.025S.The time domain waveform is mapped into corresponding spectrogram. Then the spectrogram is divided into size of 8×8 small spectrograms. After reading for this, 1000 pieces of small spectrograms are randomly chosen in each sample spectrogram, so the total small spectrograms which are used for training dictionary is5000. Fig.1 (a) describes the redundant dictionary which is trained by K-SVD algorithm. The experiment above is designed for clean speech signal. Another experiment for corrupted speech signal is conducted. The training samples used above are added the White Gaussian noise σ =5, and these corrupted signals are taken for training the K-SVD dictionary. Fig.1 (b) shows the redundant dictionary which is trained by corrupted speech signals. As compared with Fig.1 (a) and Fig. 1(b), the structures of the two redundant dictionaries are similar. So it can be concluded that the K-SVD dictionary learning process has a noise rejection capability. So that K-SVD algorithm has relatively loose constraints for choosing the training samples.



(a)

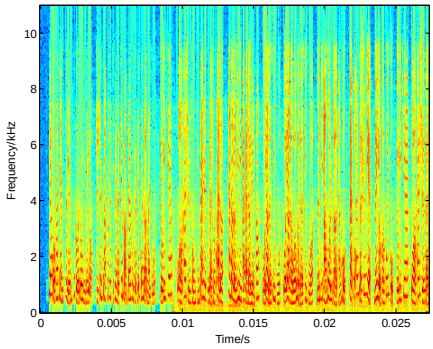
(b)

Figure 1. The Dictionaries trained by different samples,(a)This is the dictionary trained by clean samples,(b)This is the dictionaries trained by corrupted samples

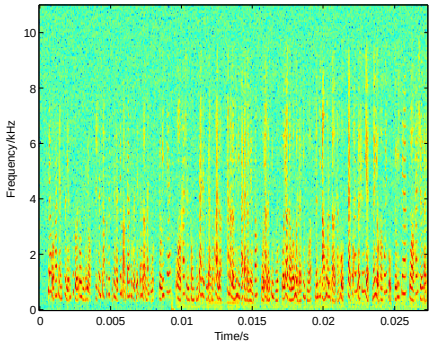
From the experiment above, it demonstrated that K-SVD algorithm has the ability to represent the useful speech signal, and the noise signal which does not have the speech structure cannot be sparse represented. So the speech denoising method proposed in this paper can utilize this characteristic of K-SVD algorithm. Notice that, the useful speech signal and the noise signal can be separated through the sparse decomposition. With this sparse decomposition the noise can be removed ultimately.

B. Contrast analysis under white noise environment

In order to test the behavior of the K-SVD denoising algorithm that uses the adaptive dictionary, here, the speech denoising experiment do for the signal with white noise is proposed. The sampling frequency of the test speech samples is 8 KHZ and frame length is 128 point. The length of each section is 0.025s and the frame range is set to 25%. Meanwhile, the time domain speech signal should be mapped into time-frequency spectrogram signal, and the spectrogram size of each signal sample is 256×256. In this experiment, in order to reduce the calculation quantity and be suitable for K-SVD algorithm, each spectrogram of 256×256 should be divided into 8×8 small spectrograms randomly (with overlaps). Through the sparse decomposition algorithm, the noise signal is removed when the small spectrogram is reconstructed. Consequently, the purpose of the whole spectrogram denoising is realized. Meanwhile, redundant DCT dictionary denoising algorithm and redundant Gabor dictionary denoising algorithm are applied for contrast. The dictionary size of redundant DCT dictionary, redundant Gabor dictionary and redundant K-SVD dictionary are all set for 64 ×256. K-SVD dictionary is trained on a data-set of 5000 small spectrograms which are randomly chosen. The termination condition is that the iterative executes 200 times. Finally, the redundant K-SVD dictionary can be obtained.



(a)

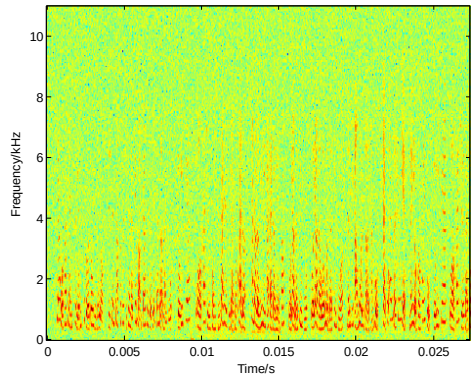


(b)

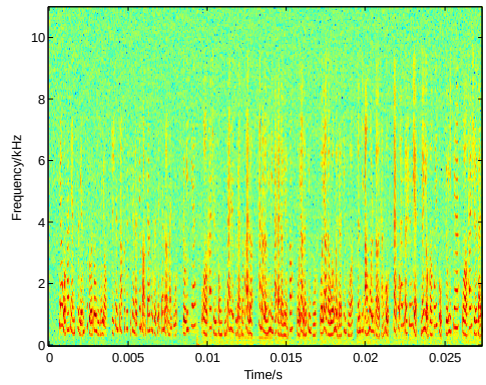
Figure 2. The test spectrogram,(a)this is the Relatively Clean Spectrogram,(b)this is the corrupted spectrogram added with noise(σ =25)

Fig.2 (a) describes a spectrogram for a relatively clean test signal, and the signal length is 25ms and the SNR is 35 db. The Gaussian White noise with variance of 25 is added into the relatively clean signal, and the spectrogram for corrupted signal is showed in Fig.2 (b).

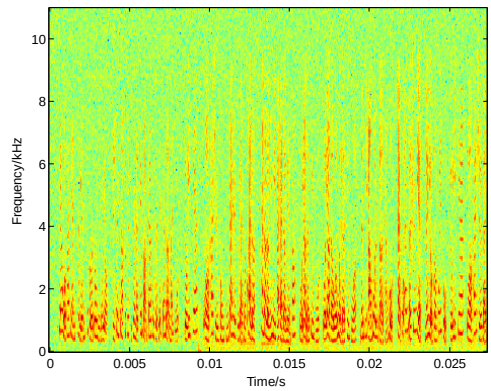
The reconstructed spectrograms are respectively presented in Fig.3. It can be seen that all the signal quality after these denoising algorithm can be improved.



(a)



(b)



(c)

Figure 3. The spectrograms after denoising,(a)this is the reconstructed Spectrogram (Gabor Denoising),(b)this is the reconstructed Spectrogram (DCT Denoising),(c) this is the reconstructed Spectrogram (K-SVD Denoising)

However, from the figures above, it can be obviously found that the method of redundant K-SVD sparse representation has the best performance compared with the others. As a result, this presented denoising algorithm not only gets a good visual effect, but also obtains a fine auditory effect. This is because the useful signal part and the noise signal part are separated qualitatively by the redundant K-SVD sparse representation algorithm. Ultimately, the useful signal can be completely extracted from the corrupted signal. Therefore, the denoising method of redundant K-SVD sparse representation is undisputed superior to the others.

IV. CONCLUSION

This work has presented a simple method for speech denoising, leading to a novel denoising performance. According to the theory of speech signal redundant sparse representation, a method for speech signal denoising which is based on redundant K-SVD dictionary is proposed. As the spectrogram can ordinary synthesize the characteristics of time domain waveform and spectrum, therefore, the spectrogram is selected for sparse representation. Significantly, this algorithm can also improve the denoising process speed. This denoising method firstly uses K-SVD algorithm to train the initial dictionary. Synchronously, OMP algorithm is applied to do sparse decomposition for the corrupted signal over the redundant dictionary. Finally, the decomposed sparse coefficients are used to reconstruct the speech signal. After all this steps, the purpose of separating the speech and noise can be achieved. Moreover, the experimental results have proved the validity of this method. In this research, the content of the dictionary is of prime importance for the denoising process. The redundant K-SVD dictionary is obtained by training on the speech signals, so the dictionary can reflects the characteristics of signal structure well. So that compared to the other speech denoising methods, this method shows a more superior performance whether in objective index or subjective quality. Although this method has effectively improved the speech signal quality to a certain extent, but there is a large research space that do not be explored in this work. For example, the research of a more complex model that uses several dictionaries switched by signal content, the research of the optimization algorithm for greed algorithm when training the dictionary, the research of a redundant dictionary and more. In a word, all of these are the research direction.

ACKNOWLEDGMENTS

This research is supported by: National Natural Science Foundations of China (Grant No.61372146), the National Natural Science Foundation of China (Grant no. 61373098), the College student practice innovation project of Jiangsu province (No:201311054001Y) , Innovative team foundation of Suzhou vocational university and the Innovative Plan Project for Graduate students of Jiangsu province(No.CXZZ13_0812)and the suzhou vocational university Young Investigator research grant(No. 2013SZDQ03).

REFERENCES

[1] Loizou P C, “Speech Denoising: Theory and practice,” CRC Press, USA, 2007.

[2] Boll S F, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. On Acoustics, Speech, and Signal Processing, vol. 27, no. 2, pp. 113-120, 1979.

[3] Xiaosheng Yu, “An improved total variation model with adaptive local constraints and its applications to image denoising, deblurring and inpainting,” JDCTA: International Journal of Digital Content Technology and its Applications, vol. 5, no. 12, pp. 170-177, 2011.

[4] Lei Zhang, Quan Pan, “On the determination of threshold in threshold based denoising by wavelet transformation,” Acta Electronica Sinica, vol. 29, no. 3, pp. 400-403, 2001.

[5] B.Yegnanarayana, Carlos Avendano, Hynek Hermansky, P. Satyanarayana Murthy, “Speech denoising using linear prediction residual,” Speech Communication, vol. 28, pp. 25-42,1999.

[6] Gowreesunker B V, Tewfik A H, “Learning sparse representation using iterative subspace identification,” IEEE Transactions on Signal Processing, vol.58, no. 6, pp. 3055-3065, 2010.

[7] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” Biometrika, vol. 81, no. 3, pp. 425–455, 1994.

[8] Wen Jin, Michael S. Scordilis, “Speech denoising by residual domain constrained optimization,” Speech Communication, vol. 48, pp. 1349-1364, 2006.

[9] Tantibundhit C, Pernkopf F, Kubin G. “Joint time-frequency segmentation algorithm for transient speech decomposition and speech denoising,” IEEE Transaction on Audio, Speech, and Language Processing, vol. 81, no. 6, pp. 1417–1428, 2010.

[10] Chen SS, Donoho D L, Saunders M A, “Atomic decomposition by basis pursuit,” SIAM Review, vol. 43, no.1, pp. 129–159, 2001.

[11] M. S. Lewicki and T. J. Sejnowski, “Learning redundant representations,” Neur. Comput., vol. 12, pp. 337–365, 2000.

[12] Yao zhang, “Hybrid recommendation method IN sparse datasets: Combining content analysis and collaborative filtering,” JDCTA: International Journal of Digital Content Technology and its Applications, vol. 6, no. 10, pp. 52-60, 2012.