ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SPEECH SIGNAL COMPRESSED SENSING BASED ON K-SVD ADAPTIVE DICTIONARY

[1,2]YAN ZHOU,[1]HEMING ZHAO

[1] School of Electronics and Information Engineering, Soochow University, China

[2] Department of Electronics and Information Engineering, Suzhou Vocational University, China

E-mail: zhyan@jssvc.edu..cn

ABSTRACT

This paper proposes a novel and successful method for speech signal compressed sensing based on K-Singular Value Decomposition (K-SVD) algorithm. K-SVD is an iterative method that alternates between sparse representation of the train samples based on the current dictionary and a process of updating the dictionary atoms to better fit the speech data. The presented K-SVD algorithm is applied here for training an adaptive overcomplete dictionary which can best suit a set of given speech signals. The sparse coefficients can be obtained by conducting the Orthogonal Matching Pursuit(OMP)sparse decomposition algorithm. At last, the original signal can be reconstructed by exploiting reconstruction algorithm. The experimental results show that, compared with the traditional basis function dictionary, the proposed method has a stronger adaptability which can be availably used for speech signal sparse representiation. Moreover, the compressed sensing based on K-SVD algorithm achieves higher reconstruction accuracy and better denoising efficiency.

Keywords: *Speech Signal, Compressed Sensing, K-SVD Algorithm, Overcomplete Dictionary, Sparse Decomposition, Signal Reconstruction*

1. INTRODUCTION

The Compressed Sensing (CS) theorem[1-3] specifies that, employing some nonadaptive linear projections that preserve the structure of the signal, the original signal then can be reconstructed from these projections by using an optimization process. In recent years, extensive research of compressed sensing for speech signal has established a solid research foundation [4-9]. The application premise of CS theorem is that the speech signal must be K-sparse. Extraction of the sparsest representation is a hard problem that has been extensively investigated in the past few years. In1993, the sparse decomposition method based on overcomplete dictionary was proposed by Zhang and Mallat [10]. An overcomplete dictionary that leads to sparse representations can either be chosen as a prespecified set of functions or designed by adapting its content to fit a given set of signal. Choosing a prespecified transform matrix is appealing because it is simpler. Also, in many cases it leads to simple and fast algorithms for the evaluation of the sparse representation. This is indeed the case for overcomplete wavelets, curvelets, contourlets, steerable wavelet filters, short-time Fourier transforms, and more[11-12]. The learning dictionaries in applications depend on how suitable they are to sparsely describe the signals in question. Multiscale analysis with oriented basis functions and a shift-invariant property are guidelines in such constructions. The K-SVD is an iterative method that alternates between sparse coding of the examples based on the current dictionary and an update process for the dictionary atoms so as to better fit the data. So the K-SVD algorithm is a novel method for adapting dictionary to represent signal sparsely and be widely used [13-18]. Because the K-SVD algorithm is flexible and can work with any pursuit method, in this paper, the K-SVD algorithm is proposed as a novel algorithm for training adaptive dictionaries in order to achieve sparse signal representations.

The compressed sensing model of speech signal includes three aspects as been shown in Fig.1. The first part is the sparse representation of speech signal, it consists of overcomplete dictionary designing and sparse decomposition algorithm choosing. In this paper, the K-SVD algorithm is proposed to train the overcomplete dictionary, and the OMP algorithm is introduced to sparse decompose speech signal. The second part of this model is the compressed transmission. In this part,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

by using linear transformation, the observation vector dimension can be dropped, and it is far less than the original signal dimension. The third part is the reconstruction of the speech signal. Since the observation vector and the observation matrix can be calculated, the essence of the reconstruction is using the greedy algorithm or base tracking algorithm to match the original speech signal. This paper mainly discusses the first part of this model and the application of it.

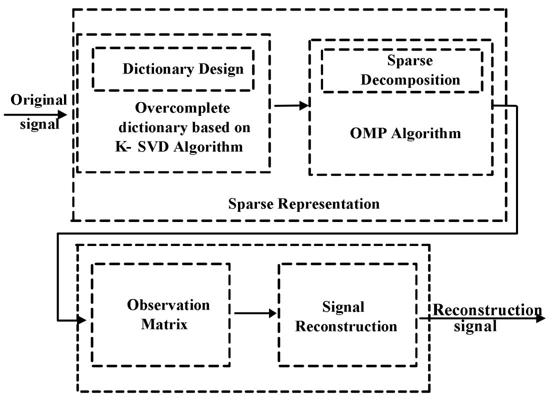


Fig.1: Compressed Sensing Model for Speech Signal

To summarize, the novelty of this paper includes the method for speech signal compressed sensing based on K-SVD algorithm.The K-SVD algorithm is described in Section II, along with the analysis of initialization of dictionary and the learning rule in detail. Also novel in this work is the idea to train dictionaries for the denoising task, rather than using prechosen ones. Dictionary training is done by using the corrupted speech directly, and followed by dictionary updating. This is described in Section III. In Section III, it also shows the experimental results that demonstrate the effectiveness of this algorithm. Section IV summarizes this paper and puts forward the research direction.

2. SPARSE REPRESENTATION ALGORITHM

In this paper, a novel route of K - SVD algorithm for designing dictionary based on learning is considered. The goal is to search the dictionary that yields sparse representations for the training signals. By this method, the achievable dictionary can be used for special classes of signals adaptively because the K-SVD algorithm can extract the inherent characteristics of train signals. Commonly, it is believed that such dictionaries have the potential to outperform the used predetermined dictionaries.

2.1. Initialization of Basis Function

The initialization of basis function is actually referred to as assignment the initial dictionary. There are two ways; one is arranging the train sample directly as to be the initial dictionary. The other is choosing the properly sparse transform matrix as the initial dictionary. Since the Gabor function is generated by Gaussian function after complex sine modulate, and it has the strong ability to express the point of singular signal. So in this paper, the Gabor atom is used for the sparse transform matrix as for the initial dictionary.

2.2. Learning Rule

Speech signal is a kind of signal with wide change range of time and frequency. However, the overcomplete dictionary trained by K-SVD algorithm can better conform to the speech signal structure and content. Consequently, according to the characteristics of the test signal, the S atoms of best linear combination to represent a signal sparsely can be adaptively find in the overcomplete dictionary. The K-SVD is an iterative method that alternates between sparse coding of the examples based on the current dictionary and an update process for the dictionary atoms so as to better fit the data. The update of the dictionary columns is done jointly with an update of the sparse representation coefficients related to it, resulting in accelerated convergence.

Firstly, the initial dictionary $D = \{d_j\}_{j=1}^{K}$ must be assumed as fixed, aligning with this, the speech signal $Y = \{y_i\}_{i=1}^{N}$ should be decomposed over this initial dictionary by using OMP algorithm. Since the iterative termination condition is decided by the decomposition residual.Consequently, the sparse matrix $X = \{x_i\}_{i=1}^{N}$ can be got by approximating the following formula.

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

Here, the D denotes the overcomplete dictionary, the X denotes the sparse matrix, the Y denotes the train signal, and the ε denotes the match error

The next step aims to update the redundant dictionary D where it is necessary to fix the sparse matrix X which has been trained at the last step. For the size of K-SVD dictionary is n×k, it is an iterative process that the dictionary is updated column by column, and for each column k=1,2,...,K

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

in dictionary D. Meanwhile, setting the overall representation error matrix as E_k, E_k is usually calculated in all samples which have been removed the composition of atom d_k. Notice that the error E_k can be calculated as:

$$E_k = Y - \sum_{k=j} d_k x_T^k \tag{2}$$

Here, X_T^k is the k-th line in spars matrix X of d_k. So this can be expressed as:

$$\|Y - DX\|_F^2 = \left\| Y - \sum_{j=1}^{K} d_j x_T^j \right\|_2^2$$
$$= \left\| \left(Y - \sum_{j \neq k} d_j x_T^j \right) - d_j x_T^k \right\|_F^2 \tag{3}$$
$$= \left\| E_k - d_j x_T^k \right\|_F^2$$

In order to ensure the convergence of this algorithm, get rid of all the zero values in x_T^k and only retain the non-zero values before doing the SVD update.

Assuming E_R^k is the error of atom d_k, it can be expressed as $E_R^k = E_k \Omega_k$, where, Ω_k is the matrix $N \times |w_k|$, $w_k = \{ i \,|\, 1 \le i \le N, x_T^k \neq 0 \}$, w_k is the index group for all the samples which been decomposed by using atom d_k, let $x_R^k = x_T^k \Omega_k$, here, x_R^k expresses the value of line vector which has been removed 0 item, so the (3) can be turned into following calculation formula:

$$\left\| (E_k - d_k x_T^k) \Omega_k \right\|_F^2 = \left\| E_k \Omega_k - d_k x_T^k \Omega_k \right\|_F^2$$
$$= \left\| E_R^k - d_k x_R^k \right\|_F^2 \tag{4}$$

Then let E_R^k apply SVD decomposition, the calculation formula is defined as:

$$E_R^k = U \Delta V^T \tag{5}$$

Thus, the first column of U is the updated atom $\tilde{d}_k$, the first column of V multiplied by $\Delta(1,1)$ is the updated line vector x_R^k.

Taken these steps repeatedly, then the updated dictionary can be obtained until it reaches iterative termination condition. When the iterative stops, the updated dictionary is fixed again and readying for doing sparse decomposition. The adaptive redundant K-SVD dictionary can be gotten by operating alternately. After acquiring the redundant dictionary, the sparse decomposition algorithm is employed to obtain the sparse coefficients.

3. EXPERIMENTAL RESULTS AND ANALYSIS

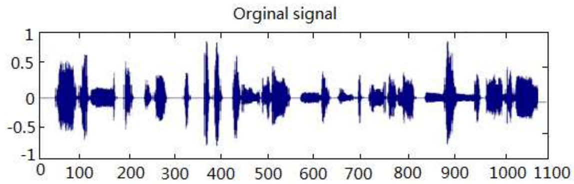
3.1. Analysis of K-SVD Overcomplete Sparse Representation

In order to verify the K-SVD algorithm on speech signals, the sparse representation experiment is executed in this part. In addition, an overcomplete separable version of the DCT dictionary by sampling the cosine wave in different frequencies is built to compare with K-SVD dictionary. At first, it is needed to prepare the training data to carry out the overcomplete dictionary. The 20 section women speech signals are chosen as the training sample which is taken randomly from the database of TIMIT. Moreover, the test signal is also randomly chosen in TIMIT, that is presented in Fig.2(a). All of these experiment performed with the signal length is 1024 dot, sampling frequency is 8Khz, frame length is128 and the window length is 256. Then the train data is constructed as a set of 2000 section speech signal which is taken from the 20section train samples. The K-SVD algorithm is applied to train a dictionary of size 64×256. The DCT dictionary size also set as 64 × 256. After learned the overcomplete dictionary, the coefficients can be computed by using OMP. Because OMP algorithm is simplicity and fast execution, it is concentrated on. The sparse coefficients performed on K-SVD dictionary and DCT dictionary is presented respectively in Fig.2(b) and Fig.2(c).

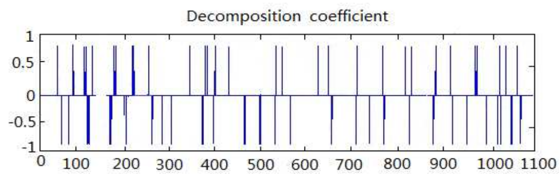
From the figures, it is denoted that sparse decomposition based on K-SVD dictionary can obtain fewer coefficients than DCT dictionary to represent original signal. So it can be concluded that speech signal over the K-SVD dictionary is more sparse, however, speech on the DCT dictionary is approximately sparse.

Journal of Theoretical and Applied Information Technology
20th February 2013. Vol. 48 No.2

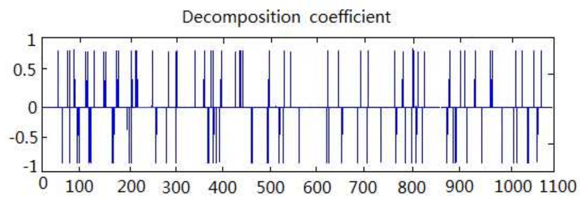
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



(A) Test Speech Signal



(B)Sparse Coefficients Based On K-SVD Dictionary



(C)Sparse Coefficients Based On DCT Dictionary

Fig.2: Sparse Representation For Speech Signal

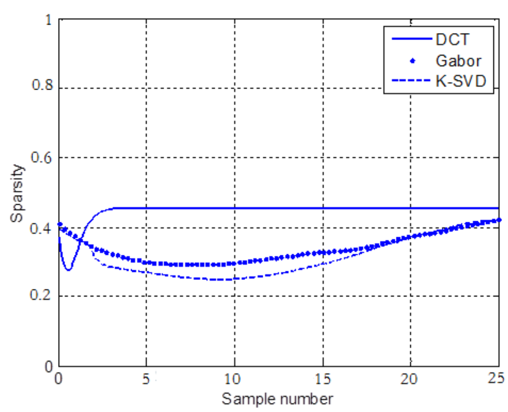


Fig.3: The Comparison Of Sparse Representation

Sparse degree refers to the non-zero vector in the sparse coefficient matrix when the algorithm reaches termination condition.Sparse degree is an important performance index which is used to measure the effectiveness of sparse representation. In order to deeply validate the sparse representation effectiveness of K-SVD dictionary, the sparse degree comparison experiment is performed. The train data and test data is the same as the experiment above, in order to increased the relative property, the Gabor overcomplete dictionary is employed for comparison. The dictionary size of all the experiments are the same. The comparison result is described in Fig.3.

As can be seen from the Fig.3, for all the samples, the average signal sparse degree based on K-SVD overcomplete dictionary is 0.3023, however, Gabor overcomplete dictionary performs 0.3452 and DCT overcomplete dictionary performs 0.4021. So it is considered that the sparse decomposition based on K-SVD has the best effect and needs the minimum storage space.

3.2. Analysis of Reconstruction Signal

The experiment above has discussed the speech sparse decomposition based on K-SVD algorithm. In this part, the reconstruction experiment is carried out. It mainly studies the influence of sparse decomposition for the signal reconstruction. According to the compressed sensing theory, it is considered that the observation vector can be obtained from the Linear Projection over the speech signal by using an observation matrix which is not related to the base function matrix. The dimension of the observation vector is less than the original signal, thus the signal can be compressed. When the compressed signal transported to the receiver, the signal can be reconstructed by applying reconstruction algorithm to approximate the original signal. This experiment aims to compare the reconstruction results with different sparse decomposition algorithms.

In this experiment, the OMP algorithm is used for algorithm reconstruction. The gaussian random matrix is introduced to be the observation matrix for compressed sensing. The mean value is 0, and the variance is 1/n . The train data is also the same with the above experiment, however, another test signal which is also randomly taken from the TIMIT database is used here. The time domain waveform diagrams areshown in the Fig.4.

Form the figures, it can be obviously seen that, when the original signal performs sparse decomposition over K-SVD overcomplete dictionary, the reconstruction signal presents the best waveform diagram.

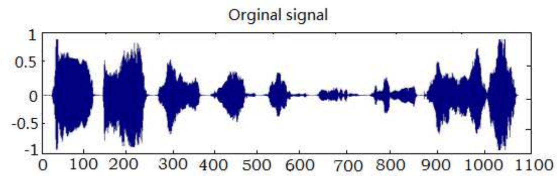
Since the signal reconstruction is critical important to the compressed sensing, here, another objective comparison experiment is conducted. The Signal-to-noise ratio(SNR) is described to evaluate the quality of reconstructed signal. The definition of SNR is as following:

$$SNB(dB) = 10\lg\left(\frac{\|Y\|^2}{\|Y - Y'\|^2}\right) \qquad (6)$$

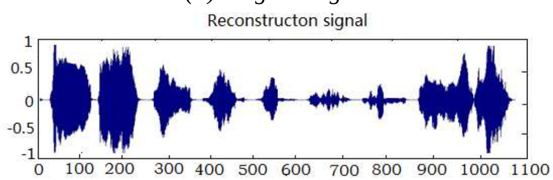
Journal of Theoretical and Applied Information Technology
20th February 2013. Vol. 48 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

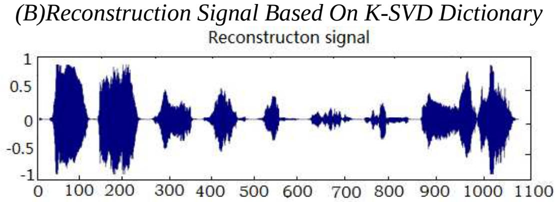
Here, Y is the original speech signal,and the $Y^{'}$ is the reconstructed signal.



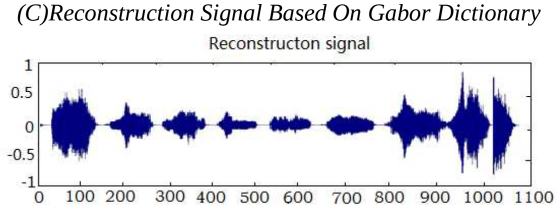
(A) Original Signal



(B)Reconstruction Signal Based On K-SVD Dictionary



(C)Reconstruction Signal Based On Gabor Dictionary



(d)Reconstruction Signal based on DCT Dictionary

Fig.4: Comparison of Reconstruction Signal Waveform

Fig.5 describes the SNR of three reconstruction signal when the sparse signal uses different observation points. As can be seen from the figure, when applying the same reconstruction condition, the signal sparse decomposition based on K-SVD dictionary gains the best reconstruction SNR result. In addition, when setting a certain reconstruction SNR, the K-SVD method needs the minimum reconstruction computational quantity. In a word, the method proposed in this paper reflects the best performance.

Finally, the method of Perception Evaluation of Speech Quality (PESQ) is introduced to test the quality of reconstruction signal.

This is a subjective evaluation method measured by Mean Opinion Score(MOS). In the experiment, ten man and ten woman testers are invited to test the quality of reconstruction signal, the testers give their scores from 5 to 1. The test headset model BOSE-QC-1 is selected. Fig.6 shows the test results of the three reconstruction signal which is based on the different sparse dictionary. The score indicates that the reconstruction signal based on K-SVD overcomplete dictionary receives the highest score. It is superior to the DCT overcomplete dictionary and Gabor overcomplete dictionary.

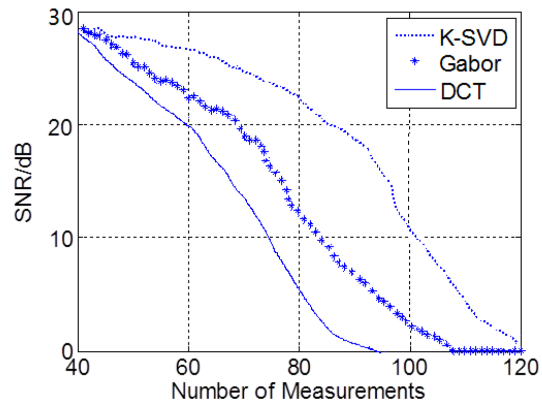


Fig.5: The SNR Comparison Of Reconstruction Signal

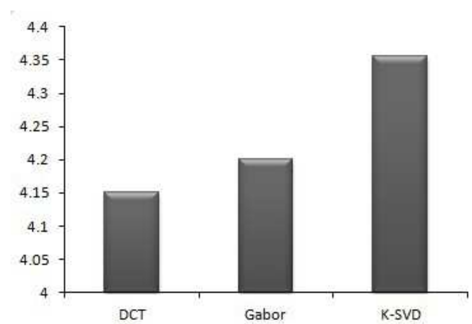
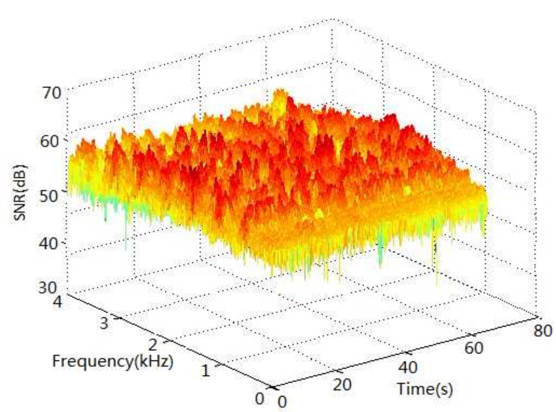


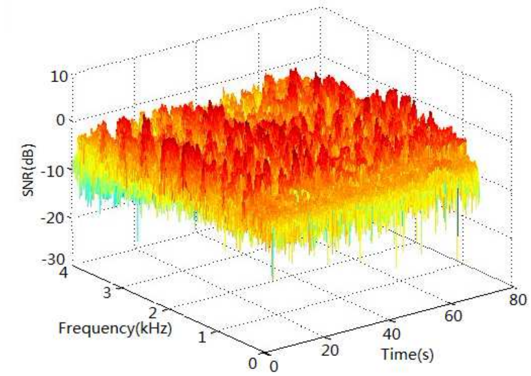
Fig.6: The Comparison Of MOS Subjective Rating Results



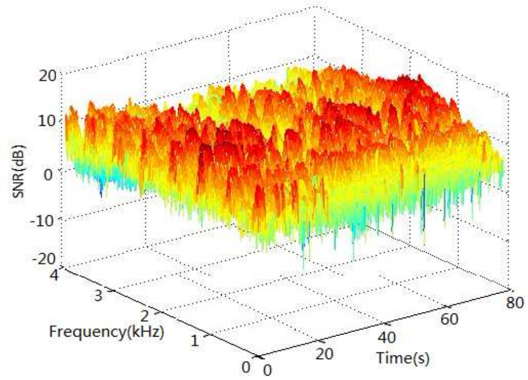
(A) The Original Speech Signal

Journal of Theoretical and Applied Information Technology
20th February 2013. Vol. 48 No.2

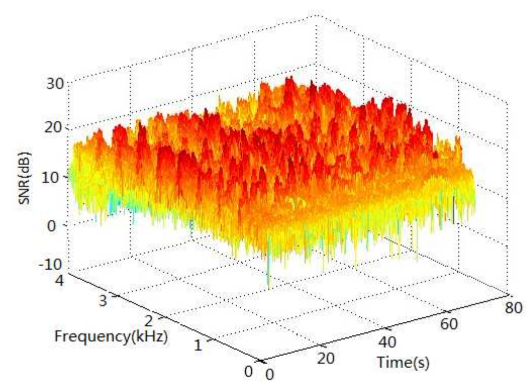
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



(B) The Corrupted Speech Signal



(C) Reconstruction Signal After Denoising Based On DCT Overcomplete Dictionary



(D) Reconstruction Signal After Denoising Based On K-SVD Overcomplete Dictionary

Fig.7:Comparison Of Signal Denoising Performance

3.3. Application Analysis of K-SVD Sparse Representation

This experiment is conducted to analyze the anti-noise performance of K-SVD algorithm, and the DCT overcomplete dictionary is employed to contrast. In the experiment, the train sample which is used above is chosen again, but the dictionary size is set as 128×896.

Fig.7 shows the three-dimensional signal among processing. The test signal is randomly chosen from the TIMIT database which is shown in Fig.7(a).The Fig.7(b) shows the corrupted signal which is added the Gaussian White Noise. In the test, the size of DCT dictionary is also set as 128×896. After ready for this, the sparse decomposition algorithm is performed. Fig.7(c) displays the reconstruction signal which is sparse decomposed by K-SVD overcomplete dictionary. Fig.7(d) displays the reconstruction signal which is sparse decomposed by DCT overcomplete dictionary. From the three dimension spectrogram, it is obviously noted that the method of K-SVD algorithm can get rid of the noise effectively. However, the DCT overcomplete dictionary based algorithm can not remove the noise effectively. On the contrary, the useful signal is submerged in the noise. Probing into the reason is that the K-SVD has the ability to reject the noise. The redundant K-SVD dictionary is obtained by training on the speech signals, so the dictionary can reflect the characteristics of signal structure well. Moreover, the speech signal can be sparse represented accurately by the K-SVD method. So that compared to the DCT based speech denoising method, this method shows a more superior performance either in objective index or subjective quality.

4. CONCLUSION

In this paper, a novel speech signal compressed sensing algorithm based on K-SVD overcomplete dictionary is proposed. The problem of generating and using K-SVD dictionary is mainly discussed. The K-SVD algorithm is for training an overcomplete dictionary that best suits a set of given speech signals. Afterwards, the OMP algorithm is applied to sparse decompose the test signal over the overcomplete dictionary, as a result, the sparse coefficients can be acquired. At last, the speech signal can be reconstructed by using the sparse coefficients. It is shown that the dictionary found by the K-SVD performs well for speech signals in applications and outperforms alternatives such as the DCT overcomplete dictionary and Garbor overcomplete dictionary. This kind of K-SVD dictionary is believed not being commonly used in speech signal processing nowadays, but it can successfully replace popular representation methods both in speech signal enhancement and in compression. The future work is required to enable such a trend for signal compressed sensing. There

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

are many possible research directions, such as the exploration of the connection between the chosen pursuit method in the K-SVD and the method used later in the application, the speech enhancement method over K-SVD dictionary for the Low SNR signal, and handling the scalability problem of the K-SVD when turning to work with larger speech data. So as to promote the development of speech signal compression sensing.

ACKNOWLEDGMENTS:

This research is supported by: National Natural Science Foundations of China (Grant No.61071215), Innovative team foundation of Soochow Vocational University (3100125) and the Innovative achievement Foundation of Soochow Vocational University (2011SZDCC06).

REFRENCES:

[1] Donoho D,"Compressed sensing", *IEEE Transactions on Information Theory*,Vol.52, No. 4, 2006, pp. 1289-1306.

[2] Donoho D,"Tsaig Y. Extensions of compressed sensing", *Signal Processing*,Vol.83, No. 3, 2006, pp. 533-548.

[3] E Candès, J Romberg and Terence Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information", *IEEE Trans. on Information Theory*, Vol.52, No. 2, 2006, pp. 489-509.

[4] Gowreesunker B V, Tewfik A H,"Learning sparse representation using iterative subspace identification", *IEEE Transactions on Signal Processing*, Vol.58, No. 6, 2010, pp. 3055-3065.

[5] Zhongke Yin, Jun Ke, "The realiztion of signal sparse decomposition based on MP using FFT", *Journal of electronics and information*,Vol.28, No. 4, 2006, pp. 614-618.

[6] Tazi, E.B.,Benabbou, A., and Harti, M.,"Robust features for noisy text-independent speaker identification using gfcc algorithm combined to vad and cmntechniques",*Journal of Theoretical and Applied Information Technology*, Vol.36, No. 2, 2012, pp. 206-216.

[7] Arthur P L, Philipos C L,"Voiced/unvoiced speech discrimination in noise using gabor atomic decomposition", *Proceedings of IEEE ICASSP*, Hong Kong, April, 2003, pp.820- 828.

[8] Rezayee,A, Gazor,S,"An adaptive KLT approach for speech enhancement", *IEEE Transactions on Speech and Audio Processing*, Vol.9, No. 2, 2001, pp. 87-95.

[9] Aharon M, Elad M, Bruckstein A, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation", *IEEE Transactions on Signal Processing*, Vol.54, No. 11, 2006, pp. 4311-4322.

[10] Mallat S, Zhang Z, "Matching pursuits with time-frequency dictionaries", *IEEE Transactions on Signal Process*, Vol.41,1993, pp. 3397-3415.

[11] Bertin N,Badeau R, and Richard G,"Blind Signal Decompositions for Automatic Transcription of Polyphonic Music: NMF and K-SVD on the Benchmark", *Proceedings of IEEE ICASSP*,April15-20, 2007,pp.65- 68.

[12] Tingting Xu, Zhen Yang, and Xi Shao, "Adaptive compressed sensing of speech signal based on data-driven dictionary", *Proceedings of 15th Asia-Pacific Conference on Communications*, Octorber 8-10,2009, pp.257-260.

[14] Giacobello D,Christensen M G,Murthi M N， Jensen S H, and Moonen M, "Retrieving sparse patterns using a compressed sensing framework: applications to speech coding based on sparse linear prediction", *IEEE Signal Processing Letters*, Vol.17, No. 1, 2010, pp. 103-106.

[15] Ravelli, E.Richard, G.,Daudet, L,"Union of MDCT Bases for Audio Coding", *IEEE Transactions on Audio, Speech, and Language Processing*,Vol.16, No. 8, 2008, pp. 1361-1372.

[16] Giacobello, D., Christensen, M.G., Murthi, M.N., Jensen, S.H., and Moonen, M, "Retrieving Sparse Patterns Using a Compressed Sensing Framework: Applications to Speech Coding Based on Sparse Linear Prediction", *IEEE Signal Processing Letters*, Vol.17, No. 1, 2010, pp. 103-106.

[17]Etemoglu, C.O, Cuperman, V,"Matching pursuits sinusoidal speech coding", *IEEE Transactions on Speech and Audio Processing*, Vol.11, No. 5, 2003, pp. 413-424.

[18]Sreenivas TV, Kleijn WB,"Compressive sensingfor sparse excited speech signals", *Proceedings of the 2009 IEEE International Conference on Acoustics， Speech and signal Processing*,Taiwan,Aprial 19-24, 2009, pp.4125-4128.