Indonesian Journal of Electrical Engineering and Computer Science
Vol. 6, No. 1, April 2017, pp. 26 ~ 35
DOI: 10.11591/ijeecs.v6.i1.pp26-35

Compressed Sensing Speech Signal Enhancement Research

Kuangfeng Ning*[1], Guojun Qin[2]
[1]School of Information Science and Engineering, Hunan International Economics University
Changsha, China, postcode: 410205
[2]School of mechanical Engineering, Hunan International Economics University
Changsha, China, postcode: 410205
*Corresponding author, e-mail: 240670821@qq.com

Abstract

The proposed Compressive sensing method is a new alternative method, it is used to eliminate noise from the input signal, and the quality of the speech signal is enhanced with fewer samples, thus it is required for the reconstruction than needed in some of the methods like Nyquist sampling theorem. The basic idea is that the speech signals are sparse in nature, and most of the noise signals are non-sparse in nature, and Compressive Sensing (CS) eliminates the non-sparse components and it reconstructs only the sparse components of the input signal. Experimental results prove that the average segmental SNR (signal to noise ratio) and PESQ (perceptual evaluation of speech quality) scores are better in the compressed domain.

Keywords*: compressed sensing, denoising, signal enhancement, effect testing*

1. Introduction

With the development of communication technology, voice communication has become a major communication medium for people to transmit information more convenient. However, the widespread nature of noise makes the voice communication quality has declined. Therefore, to reduce the noise on the performance of voice communications, improve the quality of voice communications, voice denoising for technology has become a hot research topic. S. Boll [1] in 1979 presented the classic spectral subtraction algorithm (Spectral Subtraction, SS). The algorithm assumes that short-term stationary additive noise and speech signal independent of the conditions, through the spectrum from the noisy speech signal by subtracting the estimated noise spectrum, resulting in denoised speech signal spectrum. But because of its assumption of local stability is not consistent with the actual situation, so the results are unsatisfactory, leading to larger residual musical noise and other issues. So Berouti [2] in the traditional spectral subtraction based on the increase of the size of the adjustment coefficient of the noise power spectrum and the enhanced speech power spectrum to increase the minimum limit the performance of spectral subtraction. However, due to its correction factor and the minimum value is determined based on experience, poor adaptability of the method. Y Ephraim 1984 [3] introduced the minimum mean square error to the spectral subtraction, can be part of the solution to the music noise and improve the denoising results. But the algorithm requires prior estimates because the distribution of speech spectrum, and thus larger than the calculation. P. Lochwood et al [4] on the basis of spectral subtraction, SNR of speech signals based on adaptive speech enhancement gain function, nonlinear spectral subtraction algorithm is proposed (Nonlinear Spectral Subtractor, NSS), although the algorithm to improve the voice signal to noise ratio, but the audio quality has not improved. Finally, in order to further reduce the musical noise, improving voice clarity, people continue to put forward a variety based on the traditional spectral subtraction improved algorithm [5-7], better voice quality improved. However, when the signal to noise ratio in low or non-stationary noise, the performance of the traditional spectral subtraction tends to become poor. To this end, S Kamath et al, 2002 [8] based on iterative multi-band spectrum subtraction method. The method takes into account colored noise on the speech spectrum of the inhomogeneity, the introduction of spin sub-band processing factors, while maintaining high voice quality at the same time, can effectively eliminate the noise

Received November 30, 2016; Revised February 1, 2017; Accepted February 16, 2017

pollution under the colored background noise and music noise. Also based on speech production model maximum a posteriori estimation [9], Kalman filters [10-12], it is the voice of the generation process can be equivalent to a linear time-varying filters for different types of voice using different excitation sources.

Y Ephraim et al, 1995 [13] constructed the first time in the time domain for a new speech enhancement approach (sub-space frame theory), a signal subspace based speech enhancement algorithm. The basic idea is to noisy speech signal space by some method into two orthogonal subspaces, one is the voice signal plus noise subspace, also called the signal subspace, because in this sub-space is primarily based on the signal based; the other is the noise subspace, the noise subspace contains only noise components. Therefore, the estimated clean speech can be a signal in the noise subspace removed, leaving only the signal subspace of the signal. After removing the noise subspace in the signal plus noise subspace to estimate the voice signal filtering. Y Ephraim, however initial work mainly for white noise, in order to deal with non-white noise, Mrital, 2000 [14] proposed a signal/noise KL transform, although the enhanced signals after each frame has a smaller residual noise, However, the non-frames between the stability of the residual noise disturbing. A Rezaye and S Gazor, 2001 [15] proposed an adaptive KLT approach for handling non-stationary noise, they assumed that the feature vector for the speech signal can be approximated by the non-stationary colored noise covariance matrix diagonalization. But this is a sub-optimal approach, it does not exist fast algorithm, which is the inadequacy of the method. Therefore, Y Hu et al, 2003 [16] method for Rezayee deficiencies in the signal subspace decomposition is proposed based on in the time domain and frequency domain speech enhancement for colored noise algorithm. In the same year A Lev and Y Ephraim [17] have also proposed approach for colored noise. However, the premise of the above methods are required noise covariance matrix must be full rank, which is not applicable for narrow-band noise.

The traditional Nyquist sampling theorem requires to meet that the sampling rate is not less than twice the highest frequency of the signal. With the development of signal processing technology and the surge in the amount of data processed, this sampling method has been far from the requirements to keep up with the high-speed signal processing. In 2006, Donoho proposed compressed sensing (Compressed sensing, CS) theory [18, 19]. When the signal has a sparse nature, its sparse features can been used, and the number of points, which are less than the signal sampling points, can be approximated to restore the original signal. This theory has greatly promoted the process of signal processing theory, and there are broad prospects on application. Currently, compressed sensing theory have a very good application in compressed image, converting analog information, bio-sensing, signal detection and classification, wireless sensor networks, data communications and geophysical data analysis and other fields [20,21].

2. Compressive Sensing based Speech Enhancement Scheme

With the increase in the use of mobile phone, laptops etc, communication have taken the next level and the enhancement of signals has become very important. The Nyquist Shannon sampling theorem states that for reconstruction, the input signal must have all frequencies below the Nyquist frequency, which is half the sampling frequency, so one needs to know the information of the signal in advance for reconstruction in this method, obviously large amount of samples are required in Nyquist sampling method and we have to compress them afterwards. CS method combines both compressing and sampling in a single step, and for the reconstruction a small data of non-adaptive linear measurements of the input signal is enough. As the signals can be reconstructed from a fewer measurements than the Nyquist sampling, the storage space and transmission bandwidth are reduced.

The CS method is based on the realization that the input signal x(n) has sparse representation [22,23].

$$x(n)=\Psi\,\theta(n) \quad (1)$$

Where, Ψ is $N \times N$ sparse inducing matrix.

As mentioned before, sampling and compression are performed in one step, so the entire CS method can be represented in the following matrix notation:

$$y=\Phi x(n) = \Phi\Psi\,\theta(n) \quad (2)$$

where x(n) is a N × 1 vector and the sensing matrix Φ is a M × N matrix, where M<<N, so the dimension of y(n), which is M × 1 vector is smaller than x(n). Hence sampling and compression are performed in a single step. So by varying the sensing matrix Φ, signals can be perfectly recovered. This can achieved by using a sensing matrix, which is highly incoherent to the sparsity inducing matrix Ψ, since the measurement process is non-adaptive, the sensing matrix Φ does not depend on the input signal x(n).

The last step is the recovery of the original signal, if the original signal x (n) is sparse, and the sensing matrix Φ and sparsity.

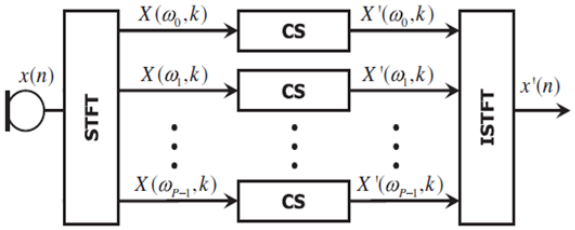

Figure 1. CS based speech enhancement scheme

Figure 1 shows the basic scheme of CS based speech enhancement. As mentioned previously, the CS method is based on the fact that speech signals are sparse in nature in the time frequency representation whereas noise signals are non-sparse. Speech signals are considered to be sparse in nature, because the speech signals are non-stationary and the speech power is not constant and it varies compared to the average power. Figure shows that the input signal x (n) is first converted to frequency domain representation through STFT block. The STFT converts the input signals in to parallel frequency bands.

Then these signals are passed through CS block, where the signals are transformed via y=Φx(n) representation, where the sensing matrix Φ can be selected from various choices are CS matrices like partial DCT or Gaussian matrices, in this project we have assumed Φ as a random Gaussian M × N matrix. As mentioned before are M<<N, the dimension of y (n) are reduced compared to x(n), hence compression is performed while the noises are suppressed. And the enhanced output signal is recovered by performing inverse STFT on all frequency bands. In this project for the CS recovery we have used the basis pursuit method. The basis pursuit de-noising formula (BPDN) states that

$$x' = \arg\min \| y - \Phi x(n) \| + \|\lambda\| \quad (3)$$

Where the de-noising process is controlled only by a single parameter λ. This parameter is called the regularization parameter and it controls the sparsity of the output denoised signal x' (n). Small value of λ makes the output less sparse and large value of λ makes the output sparser, so an optimum value of λ should provide a good trade-off between the smoothness of the reconstructed signal and the similarity to the original signal. In our project we have chosen the value for λ as 0.1, and the results demonstrate that both PESQ and SNR are better in the output signal, and both have increased as M/N and the number of STFT points increases.

3. Experimental Setup

3.1. Experimental Settings

Background noise is selected from AURORA library [24] and Noisex-92 database [25], pure voice "The birch canoe slid on the smooth planks." comes from File all_8k.wav, the sampling frequency is fs = 8kHZ.

Data selected:

all_8k_white_db -5.wav

all_8k_white_db 0.wav

all_8k_white_db 5.wav
all_8k_white_db10.wav
all_8k.wav
Using SNR

$$SNR = 10\log_{10}\left(\frac{\sum_{t=1}^{N} signal^2(t)}{\sum_{t=1}^{N} noise^2(t)}\right) \quad (4)$$

3.2. Procedures

The Figure 2 presents the procedures for the experiments of this project. The music is read by the matlab. The signal then is converted from analog to digital by the STFT matlab code, and then pass by the CS process using the l1_ls matlab code. After doing the CS, the results will be converted to analog again using the ISTFT matlab code and the SNR and PESQ are then calculated. There are three governing points in this program. First, the lambda value should be decided. Second, the program should be run using different values of the compression ratio, i.e M/N= 0.1 to 0.9 with increment of 0.1. At last, in order to get the better results, the whole program was run six times for each M/N value to calculate the average of six SNR values and six PESQ values.

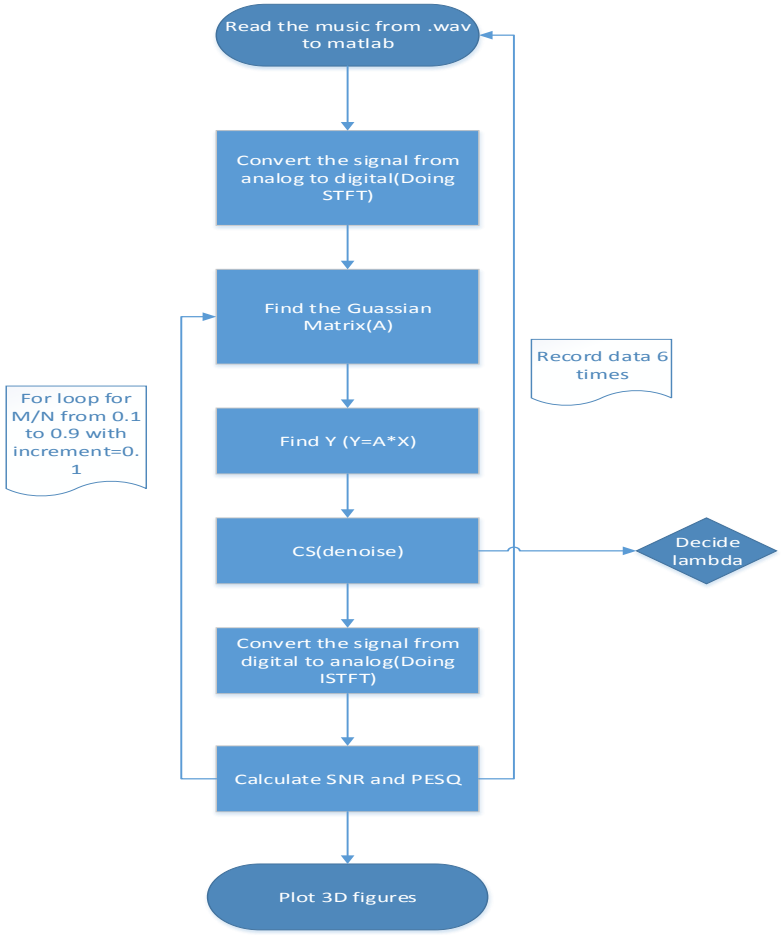

Figure 2. The flow chats of the project

Deciding Lambda value: On building up the main matlab code and running it several and several times using different values of lambda, it was clear that an acceptable and convenient value of lambda was to be taken as 0.1, this value was then fixed and considered for further measurements after that.

3.3 Performance Measures

The PESQ values and SNR values are used to evaluate the performance of the CS based algorithm. The PESQ measure is more accurate in predicting speech distortion of the processed speech whereas the SNR reflects noise suppression more accurately. Both measures give a good indication on noise suppression and speech distortion.

4. Experimental results and discussions

4.1 Data Results

On running the main matlab code 6 times for each compression ratio (M/N = 0.1:0.9), for each sub band (64, 128, 256, 512, 1024), for each signal used (all_8k_white_db -5.wav, all_8k_white_db 0.wav, all_8k_white_db 5.wav, all_8k_white_db10.wav), the following results could be obtained using the already mentioned lambda (0.1).

4.2 Plotting Data Results for Varying M/N and Sub bands

The figures 3 to 6 present the PESQ results of -5dB, 0dB, 5dB and 10dB respectively using CS process, which is presented with different M/N ratios (0.1:0.9) and different sub bands (64-1024).

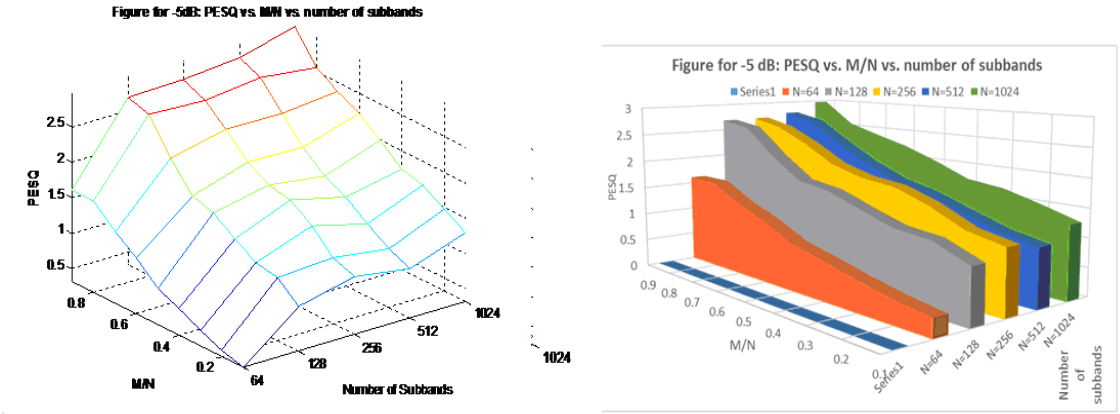

Figure 3. for -5dB: PESQ vs M/N vs number of sub bands

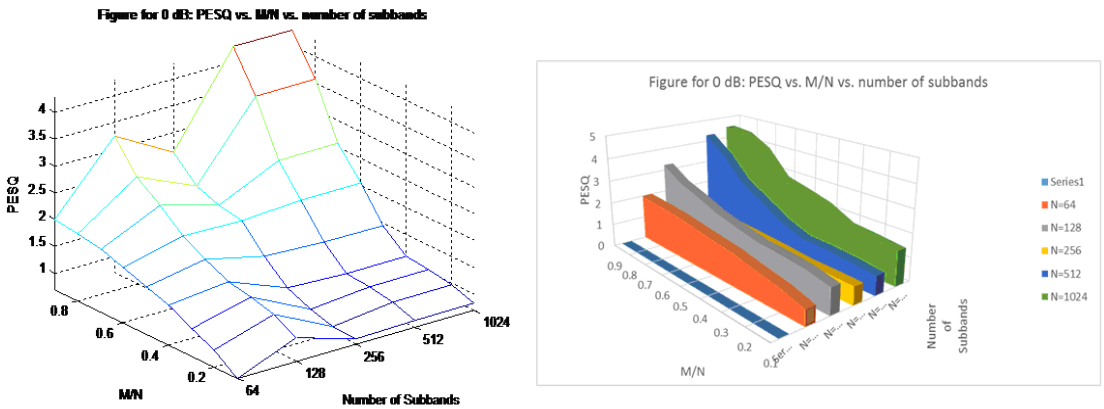

Figure 4. for 0dB: PESQ vs. M/N vs. number of sub bands

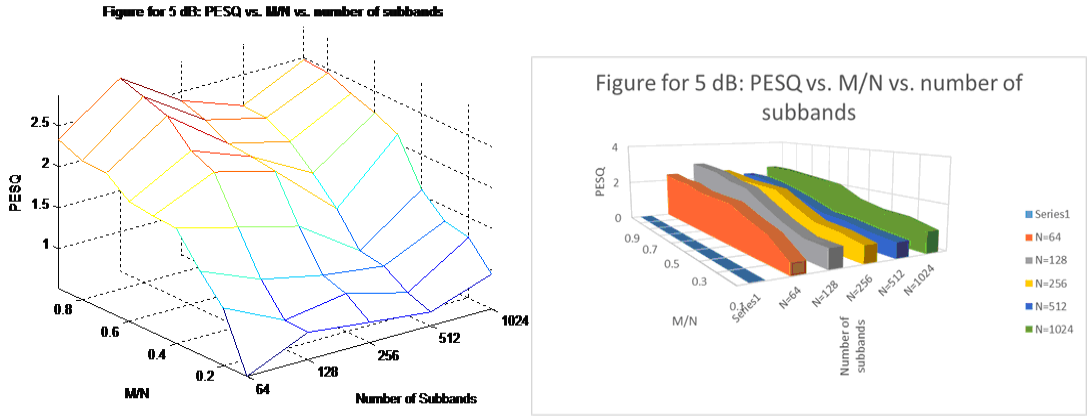

Figure 5. For 5dB: PESQ vs M/N vs number of sub bands

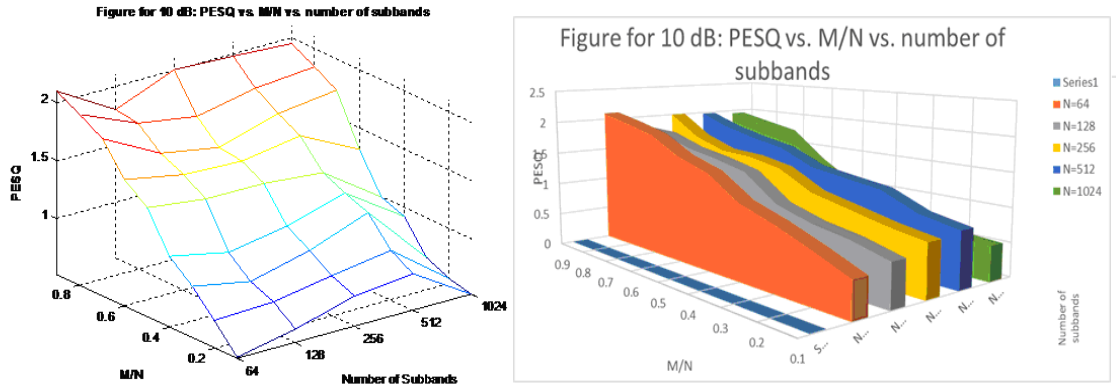

Figure 6. For 10dB: PESQ vs M/N vs number of sub bands

As it can be seen in the previous figures, The PESQ scores are generally increasing in values as the M/N increases from 0.1 to 0.9

The figures 7 to 10 present the SNR results of -5dB, 0dB, 5dB and 10dB respectively using CS process, which is presented with different M/N ratios (0.1:0.9) and different sub bands (64-1024).

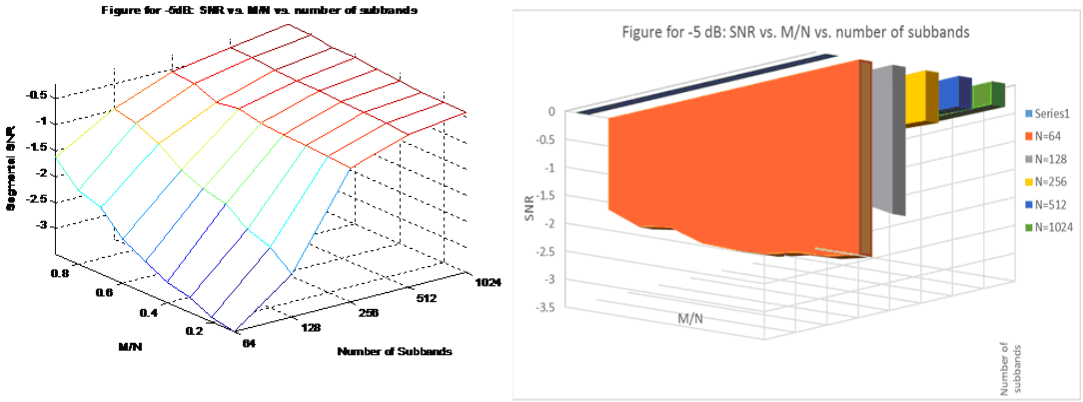

Figure 7. For -5dB: SNR vs M/N vs number of sub bands

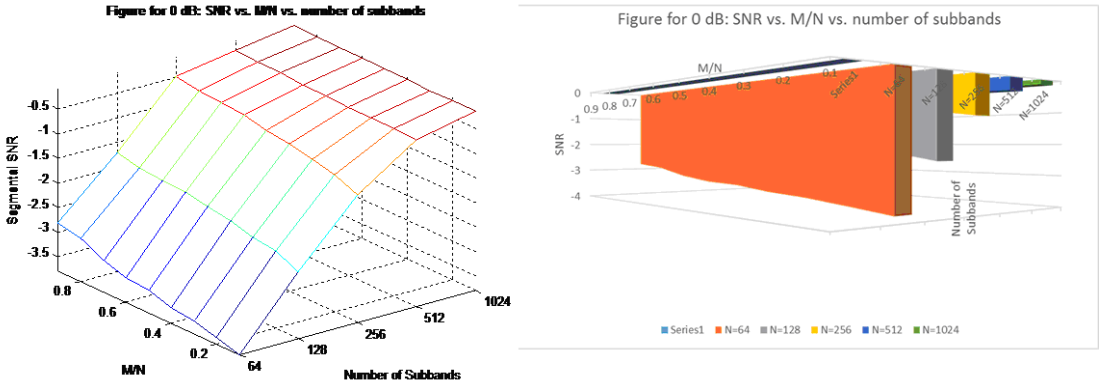

Figure 8. For 0dB: SNR vs M/N vs number of sub bands

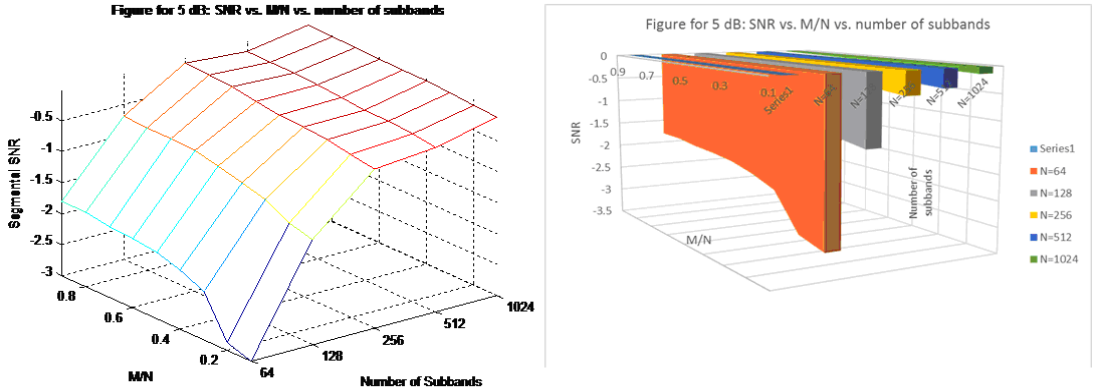

Figure 9. For 5dB: SNR vs M/N vs number of sub bands

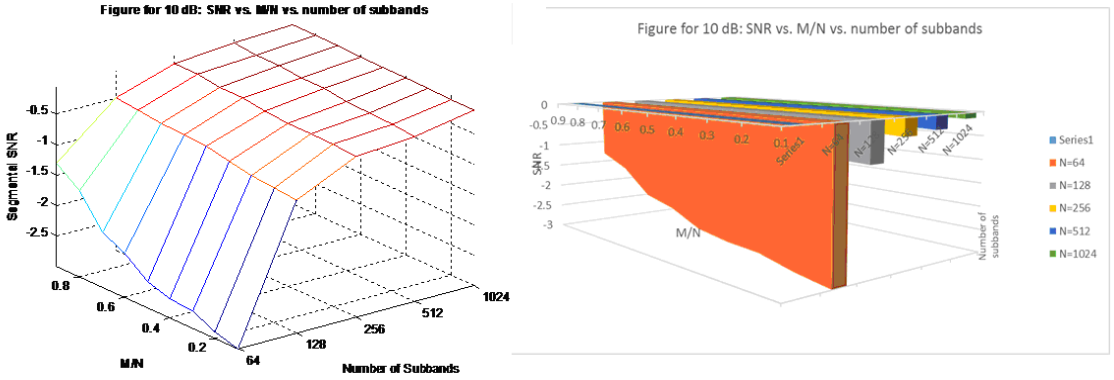

Figure 10. For 10dB: SNR vs M/N vs number of sub bands

As it can be seen in the previous figures, The SNR scores are generally increasing in values as the M/N increases from 0.1 to 0.9.

The Figures 11 to 14 present comparisons for PESQ and SNR results for -5dB, 0dB, 5dB and 10dB respectively using CS process and different sub bands (64-1024).

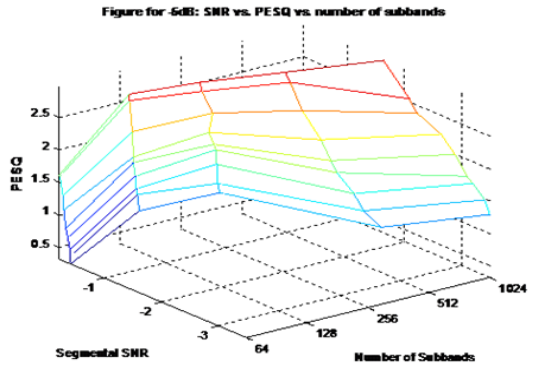

Figure 11. for -5dB: SNR vs PESQ vs number of sub bands

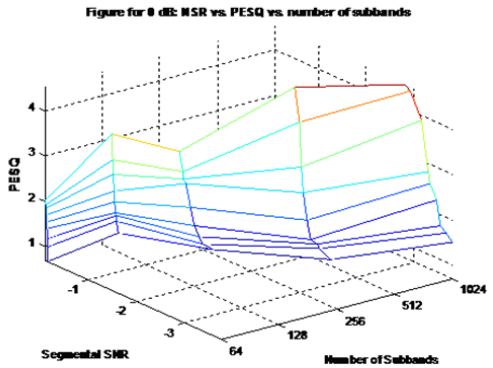

Figure 12. for 0dB: SNR vs PESQ vs number of sub bands

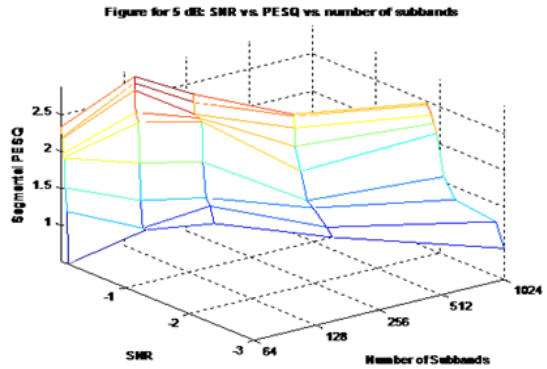

Figure 13 for 5dB: SNR vs PESQ vs number of sub bands

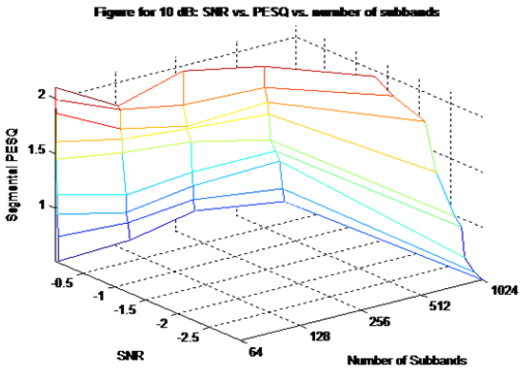

Figure 14. for 10dB: SNR vs. PESQ vs. number of sub bands

As it can be seen in the previous figures, for a higher number of STFT point's results in a slight increase of both The PESQ and the SNR scores, which is attributed to the increased sparsity of the speech observations as the number of STFT points increases.

4.3. Comparison against Other Conventional Speech Enhancement Methods

The following figure presents the performance of the proposed CS based speech enhancement is compared against two other conventional methods (OMLSA and MMSELSA methods), and as it can be seen in the figure that the CS approach is of smooth nature for different db values.

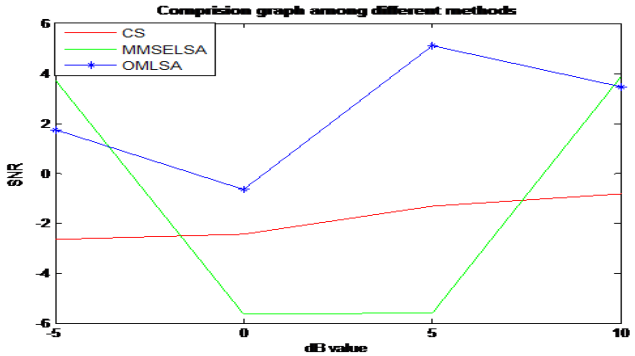

Figure 15. Comparison among other conventional speech enhancement methods

5. Conclusions and Outlook

The CS's denoising capabilities are presented by the results of PESQ and SNR. In this project, 1080 SNR values and 1080 PESQ values are recorded. In order to get more accurate results for the final plotting for the SNR and PESQ, each one needed to be measured for 9 M/N values (0.1 to 0.9) for 5 sub bands(N=64, 128,256,512 and 1024) six times each, the average number of these six times' results is calculated for plotting in matlab. The 3D plots for the SNR and PESQ are also made from Excel for good observations. The CS's denoising capabilities are presented by the results of PESQ and SNR. The results of this project proof that the CS has a great incremental contribution on the speech enhancement whereas the method results is compared to two other different methods (OMLSA and MMSELSA methods) to come out with a smoother response for different db values.

Acknowledgements

This study is sponsored by the Scientific Research Project (NO. 15C0781) of Hunan Provincial Education Department, China. This study is sponsored by the National Natural Science Foundation project (51375484) of China.

References

[1] SF Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. ASSP*, 1979; 27(2): 113-120.
[2] M. Berouti, R Schwartz, J Makhoul. *Enhancement of Speech Corrupted by Acoustic Noise*. Proceeding of 1979 IEEE, ICASSP. 1979: 208-211.
[3] Y Epharim, D Malah. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE. Trans. Acoustic, Speech Signal Processing*. 1984; 32(6): 1109-1121.
[4] P Lochwood, J Boundy. Experiments with a Nonlinear Spectral Subtractor (NSS), Hidden Markov Models and Projection, for Robust Recongnition in Cars. *Speech Commun.* 1992; 11(6): 215-228.
[5] Y Ephraim. A minimum mean square error approach for speech enhancement. *Acoustics, Speech, and Signal Processing.* 1990; 12: 829-832.
[6] Liu Zhibin, Xu Naiping. *Speech enhancement based on minimum mean-square error short-time spectral estimation and its realization*. IEEE International conference on intelligent processing system. 1997: 1794-1797.
[7] R Martin. *Speech enhancement using MMSE short time spectral estimation with Gamma distributed speech priors*. In Proc.IEEE Int.conf.Acoustics, Speech, Signal Processing. 2002; 1: 253-256.
[8] S Kamath, P Loizou. *A multi-band Spectral Subtraction Method for Enhancing Speech Corrupted by Colored Noise* . Proceedings of ICASSP. Orlando USA, IV-4164. 2002.
[9] JS Lim, AV Oppenheim. *Enhancement and Bandwidth Compression of Noisy Speech*. Proc.of the IEEE. 1979; 67(12): 1586-1604.
[10] JD Gibson, B Koo, SD Gray. Filtering of Colored Noise for Speech Enhancement and Coding. *IEEE Trans. Signal Processing*. 1991; 39: 1732-1742.
[11] WR Wu, PC Chen. Subband Kalman Filtering for Speech Enhancement. *IEEE Trans. On Circuits and Systemsli: Analog and Digital Signal Processing*. 1998; 45: 1072-1083.
[12] S Gannot, D Burshtein, E Weinstein. Iterative and sequential Kalman filter-based speech enhancement algorithms. *IEEE Trans Speech and Audio Process*. 1998; 6(4): 373-385.
[13] Y Ephraim, HLV Trees. A signal subspace approach for speech enhancement. *IEEE Transactions on Speech and Audio Processing*. 1995; 3(4): 251-266.
[14] U Mrital, N Phamdon. Signal/noise KLT based approach for enhancing speech degraded by colored noise. *IEEE Trans on Speech and Audio Processing*. 2000; 8(3): 159-167.
[15] A Rezayee, S Gazor. An adaptive KLT approach for speech enhancement. *IEEE Tram Speech Audio Processing*. 2001; 9(2): 87-95.
[16] Y Hu, P Loizou. A generalized subspace approach for enhancing speech corrupted by colored noise. *IEEE Trans on Speech and Audio Processing*. 2004; 11(4): 334-341.
[17] H Leva, Y Ephraim. Extension of the signal subspace speech enhancement approach to colored noise. *IEEE Signal Processing*. 2003; 10(4): 104-106.
[18] Donoho D. Compressed sensing. *IEEE Trans. Information Theory*. 2006; 52(4): 1289-1306.
[19] Baraniuk RG. Compressive sensing [Lecture Notes]. *IEEE Signal Processing Magazine*. 2007; 24(4): 118-121.
[20] Donoho D, Tsaig Y.Extensions of compressed sensing. *Signal Processing.* 2006; 86(3):533-548.

[21] Shi GM, Liu DH, Gao DH, Liu Z, Lin J, Wang LJ. Advances in theory and application of compressed sensin g. *Acta Electronica Sinica.* 2009; 37(5): 1070-1081. (in Chinese)
[22] Siow Yong Low, Duc Son Pham, Svetha Venkatesh, 2013. Compressive speech enhancement. Sciverse ScienceDirect Speech Communication. 2013; 55: 757–768.
[23] Dalei Wu, Wei-Ping Zhu, MNS Swamy. IET Signal Processing. 2012: 1–8, doi: 10.1049/iet-spr.2012.0192
[24] NOIZEUS: A noisy speech corpus for evaluation of speech enhancement algorithms, http://www.utdallas.edu/~loizou/speech/noizeus/ [EB/OL]
[25] Spib Noise data[EB/OL], http://spib.rice.edu/spib/select_noise.html