Indonesian Journal of Electrical Engineering and Computer Science
Vol. 3, No. 3, September 2016, pp. 489 ~ 495
DOI: 10.11591/ijeecs.v3.i2.pp489-495

Noisy Signal Processing Research based on Compressed Sensing Technology

[1] Guojun Qin, [2]Jingfang Wang
[1] School of mechanical Engineering, Hunan International Economics University, Changsha, China Changsha, 410205, China
[2] School of Information Science & Engineering, Hunan International Economics University, Changsha, Postcode: 410205, China

Abstract
Compressed sensing (CS) is a kind of sampling method based on signal sparse property, it can effectively extract the signal which was contained in the message. In this study, a new noise speech enhancement method was proposed based on CS process. Voice sparsity is used to this algorithm in the discrete fast Fourier transform (Fast Fourier transform, FFT), and observation matrix is designed in complex domain, and the noisy speech compression measurement and de-noising are made by soft threshold, and the speech signal is sparsely reconstructed and restored by separable approximation (Sparse Reconstruction by Separable Approximation, SpaRSA) algorithm, speech enhancement is improved. Experimental results show that the denoising compression reconstruction is made for the noisy signal in the algorithm, SNR margin is improved greatly, and the background noise can be more effectively suppressed.

***Keywords**: speech enhancement, compressed sensing, soft threshold, denoising, signal reconstruction*

1. Introduction

Speech is an unsteady, time-varying signal. Transmission of information by voice is one of mankind's most important and most commonly used form of information exchange. Typically, researchers understand the speech signal in relatively clean condition, and make various signal processing. But in real life, voice is inevitably affected by ambient noise and the quality and intelligibility of speech signals can been severely affected because of the presence of noise. Under the impetus of this real need, as early as the 1960s, the subject of speech enhancement is an important branch of speech signal processing, and has attracted people's attention; Since 1970s, the climax of a study had been formed and some basic results had been obtained. Currently, speech enhancement methods are spectral subtraction, Weiner filtering, Kalman filtering, and subspace enhancement, wavelet transform, and the other improved algorithm of the enhancement [1].

Speech enhancement is an important voice processing technology, and has been widely used in speech recognition, speech coding, speech synthesis and other areas. Speech enhancement aims are to extract as pure original speech from noisy speech. However, due to the noise signals are generated randomly, and there is almost impossible to completely eliminate noise. Therefore, the actual speech enhancement objectives are to improve speech clarity, improve voice quality; improve speech intelligibility [2], and to facilitate the listener to understand.

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 [3, 4]. 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

Received April 2, 2016; Revised July 25, 2016; Accepted August 10, 2016

image, converting analog information, bio-sensing, signal detection and classification, wireless sensor networks, data communications and geophysical data analysis and other fields [5,6].

Compressed sensing theory can also be applied to voice signals. Speech signal has on sparsity in some transform domain, as Fourier Transform (FFT), DCT transformation, but there is not much research, and is still at an initial stage in the current compressed sensing of the voice signal processing. Grifin has applied the CS theory in multi-channel audio signal processing [7], Giacobello has applied the CS theory in speech coding [8], Sreenivas and Kleijn considered [9] that th CS application in speech signal has not been developed, and research fields of the speech sparsity and sparse degrees is not enough depth, and, because the calculation of basic CS computation is very large, how the reconstructed perceptual characteristics of the speech signal, and how their calculative amount size is , but also these are need to focus on the practical application considerations. In China Nanjing Posts and Telecommunications University, the CS theory are combined with voice signals [10-17] to carry out a study on the observation matrix, sparse transformative matrix, voice activity detection, speech recognition systems and other aspects of anti-noise. All these indicate that CS theoryare combined with speech signal processing technology, and these are extensive research prospects.

Speech signal is a special signal, how to use the characteristics of the speech signal itself, looking for better voice signal sparse transform methods to get higher signal to noise ratio, and it would have to study the contents of this article. In this paper, a noisy speech enhancement method is researched by compressed sensing theroy, and the signal de-noising is combined with sparse signal reconstruction in the study, testing results are favorable through a variety of noise environments.

2. Signal Reconstruction Algorithm Design

In practical engineering applications, pending signals are generally contaminated to varying degrees by all kinds of noise. Noisy signal is not strictly sparse signals, but it is still a compressible signals. In existing compressed sensing theory, the fundamental basis for restoring the signal is that the signal decomposition coefficients are sparse in a transform space, but the presence of noise may destroy the signal sparsity in space. When the optimization methods are used to restore the signal, if a single sparsity constraint principle is used in the noisy signal, the original sparse signal can not been effectively restored. The other effective methods can been still used to restore the signal by compressed sensing theory, and the main difference lies in the different forms, and the objective function is optimized to set parameter in the recovery process, and the recovery effect signals are not the same with different optimization objective function.

When the signal and noise are unknown, the problem boils down to find the sparse solution with constrained quadratic programming (BSQP) problem.

$$\min \ \frac{1}{2}\| Ax-y \|_2^2 + \lambda \| x \|_1 \tag{1}$$

1. The restoration of noisy signals (1) are generalized [18, 19]:

$$\min_{x \in R^n} \phi(x) = f(x) + \lambda c(x) \qquad f: R^n \to R, c: R^n \to R \tag{2}$$

Where $f(x) = \frac{1}{2}\| Ax-y \|_2^2, c(x) = \| x \|_1$

Solution (2) is converted to the iterative (3), $\{x^t, t = 0,1,2,\cdots\}, \alpha_t > 0, \nabla f$ is gradient.

$$x^{t+1} \in \arg\min_{z} \frac{1}{2}\| z - u^t \|_2^2 + \frac{\lambda}{\alpha_t} c(z), u^t = x^t - \frac{1}{\alpha_t}\nabla f(x^t) \tag{3}$$

If $c(x)=\sum_{i=1}^{n}c_i(x_i)$ is separable, then (3) optimization into:

$$x_i^{t+1}\in\arg\min_{z}\frac{1}{2}(z-u_i^{t})^2+\frac{\lambda}{\alpha_t}c_i(z),i=1,2,\cdots,n \quad (4)$$

$$\arg\min_{z}\frac{1}{2}(z-u_i^{t})^2+\frac{\lambda}{\alpha_t}c_i(z)=soft(u_i^{t},\frac{\lambda}{\alpha_t}) \quad (5)$$

Which soft threshold shrink operations are [20]:

$$soft(u,a)=sign(u)\max\{|u|-a,0\}$$

2. α_t is selected by Barzilai-Borwein method [21] , that allow:

$$s^t=x^t-x^{t-1},r^t=\nabla f(x^t)-\nabla f(x^{t-1})$$
$$\alpha_t=\arg\min_{\alpha}\|\alpha s^t-r^t\|_2^2=\frac{(s^t)^T r^t}{(s^t)^T s^t} \quad (6)$$

When:

$$f(x)=\frac{1}{2}\|Ax-y\|_2^2$$
$$\alpha_t=\frac{\|As^t\|_2^2}{\|s^t\|_2^2},\alpha_t\in[\alpha_{\min},\alpha_{\max}]$$

3. Iteration termination condition

Given $\varepsilon(=10^{-5})>0$, if

$$\frac{|\phi(x^t)-\phi(x^{t-1})|}{|\phi(x^t)|}\le\varepsilon, or \quad \frac{\|x^t-x^{t-1}\|}{\|x^t\|}\le\varepsilon$$

Iteration is terminated.

4. Observation matrix Φ design
Sparse representation is using Fourier transform basis matrix Ψ , the observation matrix:

$\Phi=(\varphi_{ij})_{M\times N},I=\sqrt{-1}$ design

$$\varphi_{ij}=\frac{1}{\sqrt{M}}\exp(\frac{ijI\pi}{N}),i=1,2,\cdots,M;j=1,2,\cdots,N \quad (7)$$

5. y_m (i), s_m (i) and d_m (i) denote respectively the noisy speech, clean speech and additive noise in the i-th time of the m-th frame . Assuming that pure voice signal is uncorrelated with noise.

$$y_m(i)=s_m(i)+d_m(i),0\le i\le N-1,E\{s_m(i)d_m(j)\}=0,\forall i,j \quad (8)$$

FFT transform of the above equation are:

$$Y_m(\omega) = S_m(\omega) + D_m(\omega) \quad (9)$$

Wherein, Y_m (ω), S_m (ω) and D_m (ω) denote the FFT transform of the vector y_m, s_m and d_m.

By compressed sensing ($A = \Phi\Psi^T$) soft thresholding an denoising, speech spectrum:

$$\hat{S}_m(\omega) \Rightarrow \hat{s}_m(i) = F^{-1}\{\hat{S}_m(\omega)\}$$

Speech sigal is reconstructed.

3. Experimental Test Evaluation

Background noise is selected from AURORA library [22] and Noisex-92 database [23], pure voice "the birch canoe slid on the smooth planks" comes from File sp01.wav, the sampling frequency is fs = 8kHZ. In the course of the speech frame, frame size takes 25ms, ie, frame length M = [0.25fs] point, frame-shift is $\frac{M}{2}$.

Using SNR

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

It is used to quantitatively analyze the effect of de-noising algorithm. Objectively, this algorithm is improved an its performance is analyzed comprehensively from the several aspects of the speech waveform, spectrogram, SNR, segmental SNR (Time-domain segmental SNR, -10 <SNRseg <35dB), perceived voice quality assessment [24, 25] (Perceptual evaluation of speech quality, 1 <PESQ <4.5) and intelligibility fAI [26, 27] .

Experiment 1: the selected noise source is White Noise (white), and it is mixed in the voice frequency band, the algorithm can take the expected results. Figure 1 shows the contrast results of the speech enhancement before and after using the compressed sensing method.

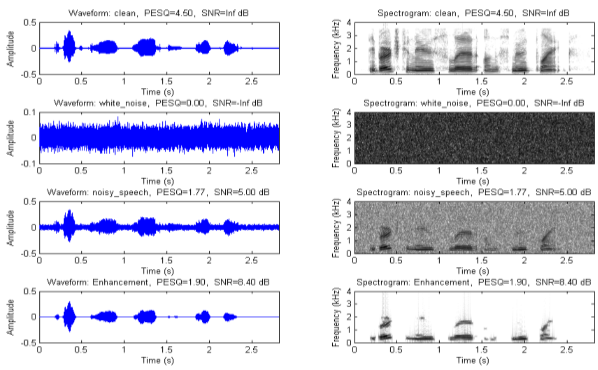

Figure 1. SNR=5dB, Results were Compared before and after the Compressed Sensing Speech Enhancement

Left part of the figure are the time domain waveforms, the right part are the spectrogram, and the original voice, white noise (white), noisy speech and enhanced speech is respectively from the top to down. Calculating SNR SNRin=5dB before noisy speech enhancement, the filtered signal to noise ratio SNRout=8.4dB. After taking the white noise, the speech enhancemen SRN efficiency:

$$\text{Efficiency} = \frac{SNR_{out} - SNR_{in}}{SNR_{in}} \times 100\%$$

The efficiency has improved 68.00% in the compressed sensing algorithm. Segmented SNR, perceived speech quality (PESQ) is respectively from -2.13dB, 1.77 up to 3.51dB, 1.90. fAI is little improvement, it is respectively 0.44,0.35 before and after the enhanced.

Experiment 2**:** Speech is enhanced by this paper compressed sensing method, and different noise in the background is loaded with the speech: White Noise (white), pink noise (pink), aviation noise (f16), factory noise (factory) and the noise of people (babble).If SNR comparison of SNR = 5dB, the contrast results in the speech waveform and spectrogram are shown in Figure 2, to examine the real-time tracking results of the algorithm.

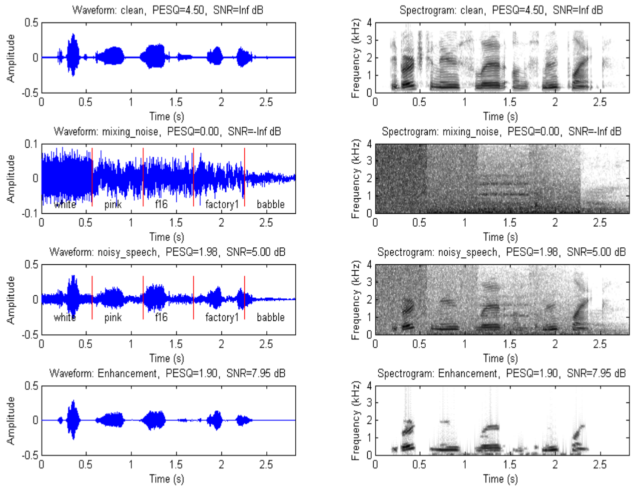

Figure 2. SNR=5dB, the Waveform and Spectrogram Comparison of Speech Enhancement Method under Same Voice and Different Noise

Left part of the figure are the time domain waveforms, the right part are the spectrogram, and the original voice, mixed noise, noisy speech and enhanced speech is respectively from the top to down. Calculating SNR SNRin=5dB before noisy speech enhancement, the filtered signal to noise ratio SNRout=7.95dB. After taking the mixed noise, the speech enhancemen SRN efficiency.

$$\text{Efficiency} = \frac{SNR_{out} - SNR_{in}}{SNR_{in}} \times 100\%$$

The efficiency has improved 59.00% in the compressed sensing algorithm. Segmented SNR is increased from -1.87dB to 2.33dB, and perceived speech quality (PESQ), fAI is respectively from 1.98dB, 0.43 up to 1.90dB, 0.31.

4. Conclusion and Outlook

A soft-threshold noisy speech enhancement method is researched in this paper, and the sparse representation of the speech signal is designe in the fast Fourier transform. The observation matrix is a complex matrix design in formula (7). Reconstruction of signal compression is combined with the approximation sparse reconstruction (Sparse Reconstruction by Separable Approximation, SpaRSA) method in the separation of the first-order norm an the second-order-norm. Experimental results show that speech is not largely improved by the PESQ, fAI evaluation in the proposed algorithm, but noisy signal compression reconstruction and signal to noise ratio can been improved by a big margin, and the background noise can been more effectively suppresse. This opens up new avenues for speech enhancement methods, and a new vision has been explored in the speech sparse sampling and reconstruction, and there is an attractive prospect.

Acknowledgements

This study is sponsored by the National Natural Science Foundation project (51375484) of China.

References

[1] Loizou PC. Speech Enhancement. Theory and Practice. USA CRC Press. 2007.
[2] Loizou PC, Gibak Kim. Reasons why Current Speech-Enhancement Algorithms do not Improve Speech Intelligibility and Suggested Solutions. *Audio, Speech, and Language Processing, IEEE Transactions on*. 2011; 19(1): 47-56
[3] Donoho D. Compressed sensing. *IEEE Trans. Information Theory.* 2006; 52(4): 1289-1306.
[4] Baraniuk RG. Compressive sensing [Lecture Notes]. *IEEE Signal Processing Magazine*. 2007; 24(4): 118-121.
[5] Donoho D, Tsaig Y. Extensions of compressed sensing. Signal Processing. 2006; 86(3): 533-548.
[6] Shi GM, Liu DH, Gao DH, Liu Z, Lin J, Wang LJ. Advances in theory and application of compressed sensing. *Acta Electronica Sinica.* 2009; 37(5): 1070-1081. (in Chinese)
[7] Gifin A, Tsakalides P. *Compressed sensing of audio signals using multiple sensors*. Proc. 16th European Signal Processing Conference (EUSIPCO' 08). Lausanne, Switzerland. 2008.
[8] Giacobello D, Christensen MG, Murthi MN, Jensen SH, Monen M. Retrieving sparse patterns using a compressed sensing framework: applications to speeeh coding based on sparse linear prediction. *Signal processing Letters, IEEE*. 2010; 17(l): 103-106.
[9] Sreenivas TV, Kl eijn W B. *Copressive sensing for sparsely excited speech signal*. Proceedlin of the 2009 IEEE International Conference on Acousties, Speech and Signal procesing, Taipei, Taiwan. 2009: 4125-4128.
[10] Guo HY, Yang Z. Compressed speech signal sensing based on approximate KL. *Journal of Electronics & Information Technology*. 2009; 31(12): 2948-2952. (in Chinese)
[11] Ye L, Guo HY, Yang Z. Research on antinoise methed of speaker recogn ition system based on compressed sensing reconstruction signal. *Signal Processing*. 2010; 26(3): 321-326. (in Chinese)
[12] Xu TT, Yang Z, Shao X. *Novel speech secure communication system based on information hiding and compressed sensing*. Fourth International Conference on Systems and Networks Communications (ICSNC 2009), Porto, Portugal. 2009: 201-206.
[13] Sun LH, Yang Z. Distributed speech cornpression and reconstruction based on compressed sensing theory. *Signal Processing.* 2010; 26(6): 824-829. (in Chinese)
[14] Ye L, Sun LH, Yang Z. Endpoint detection algorithm based on cepstral distance of compressed sensing measurem nts off speech signal. *Signal Processing.* 2011; 27(l): 67-72. (in Chinese)
[15] Ji YY, Yang Z. Compressed speech signal sensing based on antocorelative measurement. *Signal processing.* 2011; 27(2): 207-214. (in Chinese)
[16] GAO Y, WANG GM, CHEN YP, MIN G, DU J. Compressed Sensing of Speech Signals Based on Diference Transformation. *Signal processing.* 2011; 27(9): 1434-1439. (in Chinese)
[17] Zhou XX, Wang AN, Sun H, Yang HW. Speech enhancement based on compressive sensing. Tsinghua Univ Sci&Tech. 2011; 51(9): 1234-1238. (in Chinese)
[18] Wright SJ, Nowak RD, Figueiredo MAT. Sparse Reconstruction by Separable Approximation. *Signal Processing, IEEE Transactions on*. 2009; 57(7): 2479-2493.
[19] Figueiredo MAT, Nowak RD, Wright SJ. Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems. *IEEE Journal of Selected Topics in Signal Processing*. 2007; 1(4): 586-597.
[20] Donoho DL. De-Noising By Soft-Thresholding. *Information Theory, IEEE Transactions on*. 1995; 41(3): 613-627
[21] L Grippo, M Sciandrone. Nonmonotone globalization techniques for the Barzilai-Borwein method. *Computational Optimization and Applications*. 2002; 32: 143–169.

[22] NOIZEUS: A noisy speech corpus for evaluation of speech enhancement algorithms, http://www.utdallas.edu/~loizou/speech/noizeus/ [EB/OL]
[23] Spib Noise data[EB/OL], http://spib.rice.edu/spib/select_noise.html
[24] Ma J, Hu Y, Loizou P. Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. *Journal of the Acoustical Society of America.* 2009; 125(5): 3387-3405.
[25] Hu Y, Loizou P. Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on Speech and Audio Processing*. 2008; 16(1): 229-238.
[26] ITU-T Recommendation P. 862. Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs. 2000.
[27] Loizou P, Ma JF. Extending the articulation index to account for non-linear distortions introduced by noise-suppression algorithms. *Acoustical Society of America.* 2011; 130(2): 986-995.