International Journal of Multimedia and Ubiquitous Engineering
Vol.9, No.7 (2014), pp.47-60
http://dx.doi.org/10.14257/ijmue.2014.9.7.05

Study on Joint Speech Encoding Technology based on Compressed Sensing

Guangchun Gao, Lina Shang and Kai Xion

School of Information Science and Electronic Engineering, Zhejiang University City College, Hangzhou, China, 310015
seesky88@126.com, shangln@zucc.edu.cn, xiongk@zucc.edu.cn

Abstract

In this paper, a new joint speech encoding scheme based on compressed sensing was proposed. In this encoding algorithms, the compressed sensing reconstruction and PCM (Pulse Coding Modulation) were used for the speech signal encode. For the speech signal, the high frequency and low frequency coefficients can be acquired by using the wavelet transform based on the lifting scheme. For the details coefficients, using a hard threshold to remove the smaller coefficients, the high frequency coefficients are sparse, so the high frequency can be reconstructed with the compressed sensing method. Because the length of the low frequency coefficients is half of the original signal length, the PCM complex of the speech signal can be reduced. Finally, the speech can be approximately recovered with the low frequency PCM coefficients and the high frequency compressed sensing reconstructed coefficients. The experimental results demonstrate the application effectiveness for this encoding scheme in the speech processing fields.

Keywords: *Compressed Sensing, PCM, Wavelet Transform, Measurement Matrix*

1. Introduction

At present, audio coders based on sparse decompositions have already provided excellent compression, and have been used for the MPEG audio coding standards for a long time. However, the researchers still hope for a better performance speech encoding technology.

Recently, an important new theoretical advance in what has been dubbed compressive sampling or compressed sensing (CS) have been made [1-4] and has spawned a flurry of activity in research on this topic. The theory showed that a signal having a sparse or compressible representation in one basis can be recovered from projections onto a small set of measurement vectors that are incoherent with the sparsity basis. The groundbreaking work by Candes *et al.* [1] and Donoho [2] showed that such a signal can be precisely reconstructed from only a small set of random linear measurements (smaller than the Nyquist rate), implying the potential of dramatic reduction of sampling rates, power consumption and computation complexity in digital data acquisitions.

Despite the achievements have been acquired in CS theories, There are even fewer studies on the applicability of CS to audio signals, particularly on speech, music or naturally-occurring signals such as animal calls and environmental sounds. All of these signals are usually not sparse and have a large number of non-zero components in whatever basis might be used in reconstruction. The CS paradigm was used for audio compression in [5]. Relying on some classical techniques to solve the CS problem, such as basis pursuit and orthogonal matching pursuit, the method derived in [5] constructs a sparse discrete cosine transform representation of the underlying audio signal. At present, there still exists a huge gap between the CS theory and applications to audio signals [6, 7]. The researchers hope that the speech

ISSN: 1975-0080 IJMUE

encoding method based on compressed sensing can replace the traditional speech encoding methods. In this paper, we proposed the joint speech encoding method based on compressed sensing and traditional PCM speech encoding method.

In CS, it includes three main steps: sparse representation, measurement, reconstruction. The signal sparse representation is the fundamental premise of CS implementation, so the optimal sparse bases of the signals were widely researched [8-12]. For the problem of making a sparse representation of an audio signal, we introduce the DWT which can approximate smooth functions very efficiently. It can achieve arbitrary high accuracy by selecting appropriate wavelet basis, it can concentrate the large wavelet coefficients in the low frequencies, and it has a multiresolution framework and associated fast transform algorithms. The number of measurements is smaller in compressed sensing, the computational complex is lower. The measurement matrix size usually was decided by the length of the signal, so we hope that the signal can be shorten with some methods when the signal precision meet the demand of some application. For one dimension signal, it well known that the length of the low frequency signal can be shorten the half of original signal length when the signal finished the one level wavelet decomposition. Considering the signal approximation reconstruction in speech encoding, the wavelet transform based on lifting scheme was used in this paper.

Through adopting the wavelet transform, the high frequency and low frequency coefficients can be acquired. Because the high frequency coefficients usually are sparse, it can be reconstructed with CS method. In CS reconstruction, the L1-norm optimal algorithms were used. For the low frequency coefficients, which have better approximation with the original speech signal, they were encoded in PCM (Pulse Coding Modulation).

According to above analyses, for reducing the computational complex and improving the compression, we proposed the new joint speech encoding scheme based on compressed sensing theory and PCM technology. Firstly, one level wavelet decomposition of the speech signal was finished. Secondly, the high frequency coefficients were reconstructed by compressed sensing method, and the low frequency coefficients were encoded using PCM. Finally, using the inverse adaptive wavelet transform, the speech signal was reconstructed with the encoded low frequency coefficients and the reconstructed the high frequency coefficients.

2. Joint Speech Encoding Scheme

The majority of the literature on CS has been concerned with very sparse signals, and very few results have been presented that explore the performance of CS when used with signals that are not truly sparse. There are even fewer studies on the applicability of CS to audio signals, particularly on speech, music or naturally-occurring signals such as animal calls and environmental sounds. All of these signals are usually not sparse and have a large number of non-zero components in whatever basis might be used in reconstruction.

For studying on compressed sensing application in speech codec and further improving speech codec performance, the joint speech encoding scheme based on compressed sensing proposed by this paper. The codec scheme of the speech can be shown in Figure 1, which includes the compressed sensing encoding algorithms and the traditional speech encoding algorithms. In order to simplify the research work, the PCM algorithms were chosen as the traditional speech encoding algorithm.

In sender part, for the reduced the signal encoding complex, firstly, the original speech finished the one dimension wavelet transform, then the high frequency and low frequency coefficients can be acquired. For the high frequency coefficients, these values are usually close to zero. Through setting the hard threshold, the high frequency coefficients must be sparse, so these coefficients can be reconstructed by compressed sensing methods. In this

period, the high frequency coefficients were measured by the random matrix, so the measure values can be obtained in the sender. For the low frequency coefficients, according to the wavelet transform performance, it was well known that it has the better approximation of the original speech. In this paper, the low frequency coefficients were encoded by PCM method. Because the length of the low frequency coefficients was shortened, the PCM complex would be reduced.

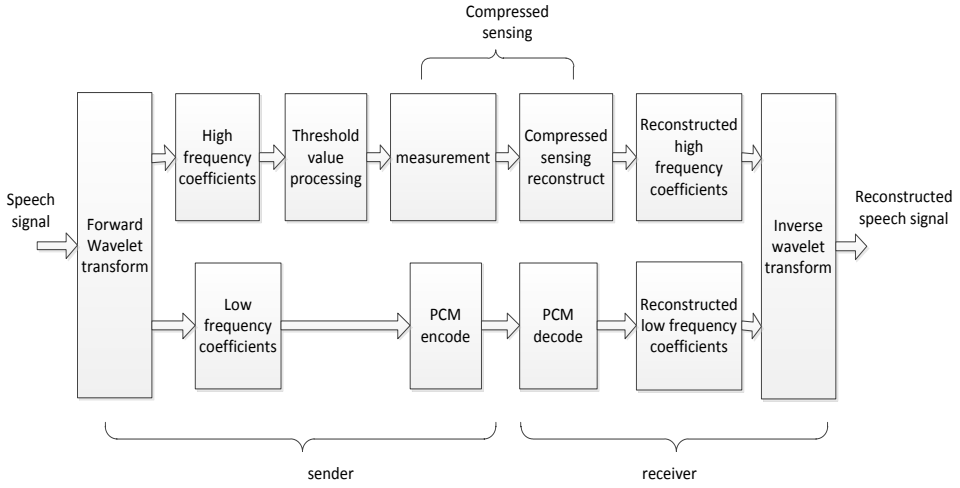

Figure 1. The Joint Speech Codec Scheme

In receiver, based on the measurement of the high frequency coefficients, the high frequency coefficients can be reconstructed by using compressed sensing algorithms. In this paper, the CVX which is a matlab-based modeling system for optimization was used for the reconstructed algorithms in compressed sensing. The reconstructed high frequency coefficients were prepared for the input coefficients of the inverse wavelet transform. The PCM code of low frequency coefficients can be decoded by using PCM decode, and then the reconstructed low frequency can be achieved in the receiver. After the low frequency and high frequency were reconstructed, the speech signal can be reconstructed with the inverse wavelet transform.

3. Compressed Sensing

During last three decades, the assessment of potential of the sustainable eco-friendly alternative sources and refinement in technology has taken place to a stage so that economical and reliable power can be produced.

The signal reconstruction system based compressed sensing was showed in Figure 2. For a signal X of size N , the nth element of the signal vector is referred to as $X(n)$, where $n = 1, \ldots, N$. Let us assume that the basis $\Psi = [\psi_1, \ldots, \psi_N]$ provides a K-sparse representation of x :

$$x = \Psi \theta \tag{1}$$

Where θ is a $N \times 1$ vector with K-nonzero elements. Many different basis expansions can achieve sparse approximations of the signal, including DCT, wavelets, and Gabor frames. In

other words, a signal does not result in exactly K-sparse representation; instead its transform coefficients decay exponentially to zero.

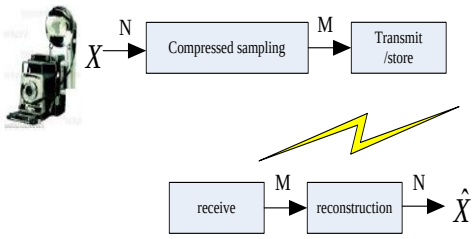

Figure 2. Signal Reconstruction based on Compressed Sensing

In the CS framework, it is assumed that the K-largest $\theta(n)$ is not measured directly. Rather, $M < N$ linear projections of the signal vector x onto another set of vectors $\Phi = [\phi_1^{'} \dots, \phi_M^{'}]$ are measured:

$$y = \Phi x = \Phi\Psi\ \theta \tag{2}$$

Where the vector y $(M \times 1)$ constitutes the compressive samples and matrix $\Phi(M \times N)$ is called the measurement matrix. Since $M \prec N$, recovery of the signal x from the compressive samples y is underdetermined; however, the additional sparsity assumption makes recovery possible.

The sparsity can be used to recover a signal that is a solution of the following minimization

$$\arg\ \min\ \|\theta\|_0 \text{ subject to } \Phi\Psi\ \theta = y \tag{3}$$

The minimization (3) is however combinatorial and thus intractable. It is relax by using ℓ^1 norm $\|\theta\|_1$ of the coefficients of x in Ψ. The recovered signal x^* is a solution of the following convex problem

$$x^* \in \arg\ \min\ \|\theta\|_1 \text{ subject to } \Phi\Psi\ \theta = y \tag{4}$$

Because M is often much smaller than N, the recovery can be view as a sort of compression. After we have solved θ from the minimization problem, we can obtain x from Equation 1[13]. The L1 optimization problem in Equation 4 can be solved with linear programming methods. The CVX software was used in this paper.

4. Wavelet Transform

Compressed sensing includes three parts: sparse representation, measurement, reconstruction. If the signal length is bigger, the measurement matrix size is bigger, so the computational complex of the signal reconstruction is higher. For reducing the size of the measurement matrix in compressed sensing, we hope that the signal length can be shortened. It is well known that the wavelet transforms can concentrate the large wavelet coefficients in the low frequencies of the signal, so the high frequency coefficients are small and close to zero. The high coefficients were sparse when they were processed by hard threshold, so the

high coefficients can be reconstructed by compressed sensing methods, and the signal length in compressed sensing is smaller than that of the original signal.

Wavelet transforms based on lifting schemes [14-16] have achieved large recognition in the last years. In general, lifting splits a signal into two sub samples, followed by at least two lifting steps, Prediction and Update, which was shown in Figure 3. The implementation process was discussed in the following text.

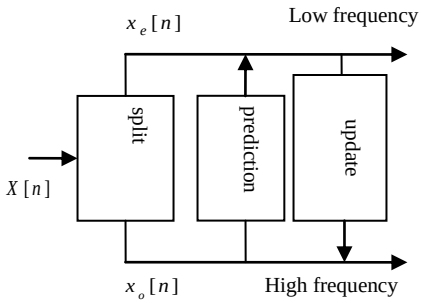

Figure 3. Prediction-then-update Scheme

4.1. The Forward Transform

To recapitulate, let us consider a simple lifting scheme with only one pair of lifting steps to go from level J+1 to level J.

Splitting: Partition the data set λ_{j+1} into two distinct data, Sets λ_j and γ_j.

Prediction: Predict the data in the set γ_j by the data set λ_j and replace γ_j by the prediction error:

$$\gamma_i - P(\lambda_j) \rightarrow \gamma_j \tag{5}$$

Update : Update the data in the set λ_j by the data in set γ_j :

$$\lambda_j + U(\gamma_j) \rightarrow \lambda_j \tag{6}$$

These steps can be repeated by iteration on the λ_j , creating a multi-level transform or multi-resolution decomposition.

4.2. The Inverse Transform

One of the great advantages of the lifting scheme realization of a wavelet transform is that it decomposes the wavelet filters into extremely simple elementary steps, and each of these steps is very easily invertible. As a result, the inverse wavelet transform can always be obtained immediately from the forward transform. The inversion rules are obvious: revert the order of the operations, invert the signs in the lifting steps, and replace the splitting step by a merging step. Thus, inverting the three step procedure above results in

Inverse update: $\lambda_j - U(\gamma_j) \rightarrow \lambda_j$ (7)

Inverse prediction: $\gamma_j + P(\lambda_j) \rightarrow \gamma_j$ (8)

Merge: $\lambda_j \cup \gamma_j \rightarrow \lambda_{j+1}$ (9)

Through the previous discussion, we know that the adaptive prediction algorithm can be reconstructed without using extra additional information. The inverse transform algorithm is the inverse process of the forward wavelet transform.

5. PCM

In order to simplify coding process, PCM was used for the speech signal encoding. Pulse code modulation (PCM) is a method of converting an analog message waveform to a digital bit stream of l's and 0's. For example, PCM is commonly used in telephone exchanges to convert an analog voice signal (300-3400 Hz) to a 64,000 bit per second data stream. Figure 4 shows the PCM generator.

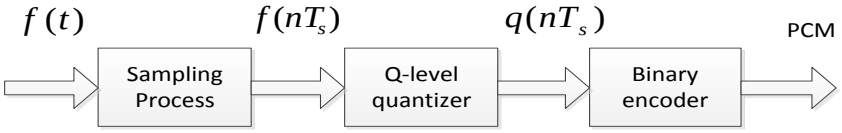

Figure 4. PCM Modulator

In the PCM generator, the quantizer compares the input $f(nT_s)$ with its fixed levels. It assigns any one of the digital level to $f(nT_s)$ that results in minimum distortion or error. The error is called quantization error, thus the output of the quantizer is a digital level called $q(nT_s)$. The quantized signal level $q(nT_s)$ is binary encoded. If $q = 2^v$, the encoder converts the input signal to v digits binary word. In this paper, the input signal of the quantizer is the speech signal, and the sampling process was omitted.

Figure 5 shows the block diagram of the PCM receiver. The receiver starts by reshaping the received pulses, and then converts the binary bits to analog. Because the permanent quantization errors were introduced during quantization at the transmitter, the original signal $f(t)$ cannot be reconstructed perfectly. The quantization error can be reduced by the increasing quantization levels v. This corresponds to the increase of bits per sample (more information). But increasing bits v increases the signaling rate and requires a large transmission bandwidth. The choice of the parameter for the number of quantization levels must be acceptable with the quantization error. Considering the simulation requirements, as the PCM generator, in PCM demodulator, the Binary encoder was considered, and D/A converter and Low-pass filter were not adopted in this paper.

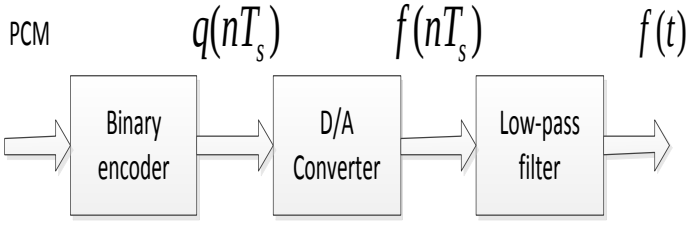

Figure 5. PCM Demodulator

6. Experimental Results

In this paper, the Figure 6 shows the speech signal which was used in experimental process. Considering the calculation complex, a part of the speech signal which was shown in Figure 7 was chosen as the test signal, and the length of the test signal is 1024.

According to the joint speech encoding scheme proposed by this paper, the test speech signal finished the wavelet transform firstly. Using the (5, 3) wavelet transform scheme in this paper, one level wavelet decomposition for the test speech signal is shown in Figure 8. The left part corresponds to the low frequency coefficients and the right part the high frequency coefficients. We notice that there are not some large high frequency coefficients. This illustrates that the wavelet coefficients have better distribution, *i.e.*, no large coefficients in the high frequencies and the energy is concentrated in the low frequency. For the joint speech encoding method, the low frequency and high frequency coefficients must be encoded separately. The following text discussed respectively the low and high frequency coefficients encoding.

For the high frequency coefficients, the compressed sensing methods were used for the signal reconstructed. As we can observe in Figure 8, the high frequency coefficients of the speech signal are not sparse in the frequency domain. Considering the signal must be sparse in compressed sensing, the high frequency coefficients are thresholded. When 0.013 was chosen as the threshold, the high frequency coefficients which were shown in Figure 9 are sparse. In compressed sensing, the CVX software was used for signal reconstruction, and then Figure 10 shows the reconstructed high coefficients. By comparing the figure 9 and figure 10, we known that the high frequency coefficients can be reconstructed perfectly.

The Figure 11 shows the low frequency coefficients, which was encoded with the PCM methods. In PCM, the numbers of quantization levels are 32, so the each quantization code can be represented by using 5 binary digits. Figure 12 presents the quantization out of the Speech signal PCM.

If the signal was encoded by PCM, the encoding quantization output was shown in figure 13. Based on the joint speech encoding scheme proposed by this paper, the signal was reconstructed, which was illustrated in Figure 14. In order to objectively evaluate the quality of reconstructed signals, the SQNR (Signal to quantization noise ratio) can be calculated by the following equation.

$$SQNR = 20 * \log\left(\frac{norm(x)}{norm(x - \tilde{x})}\right) \tag{10}$$

Where x is the original signal, and $\tilde{x}$ is the reconstructed signal. If the signal was encoded by PCM method, SQNR is 26.3538. If the signal was encoded using our methods, SQNR is 25.5858.

Otherwise, because the quantization levels is 5, each quantization value can be encoded 5 binary digits, so the signal encoded by PCM are 1024*5/8=640Byte. Because the half of the signal was encoded by PCM in our method, the PCM encoded code are 512*5/8=320Byte. In compressed sensing reconstruction processes, the measurement value is 30. If each measurement value was encoded using 8 binary, the data is 30 byte. Therefore, the encoded data is 320+30=350Byte, then the encoded data can be reduced 45.3%.

In this paper, the one level wavelet transform was finished, if the wavelet transform levels increased, the encoded files can be further reduced. Of course, the signal code qualities need to be reevaluated. The simulation results demonstrated that the joint speech encoding scheme based on compressed sensing proposed this paper can acquire better reconstruction

effectiveness and compression ration, and this speech encoding scheme can meet the demand of the speech signal reconstruction quality.

7. Conclusion

We know today that most of existing works in CS remain at the theoretical study. In particular, the implementation processes in compressed sensing have high computational complex. This paper presented the joint speech encoding scheme based on compressed sensing which can reduce computational complex and improve compression quality. The effect of the wavelet decomposition levels on the algorithms need to be further discussed in the future.

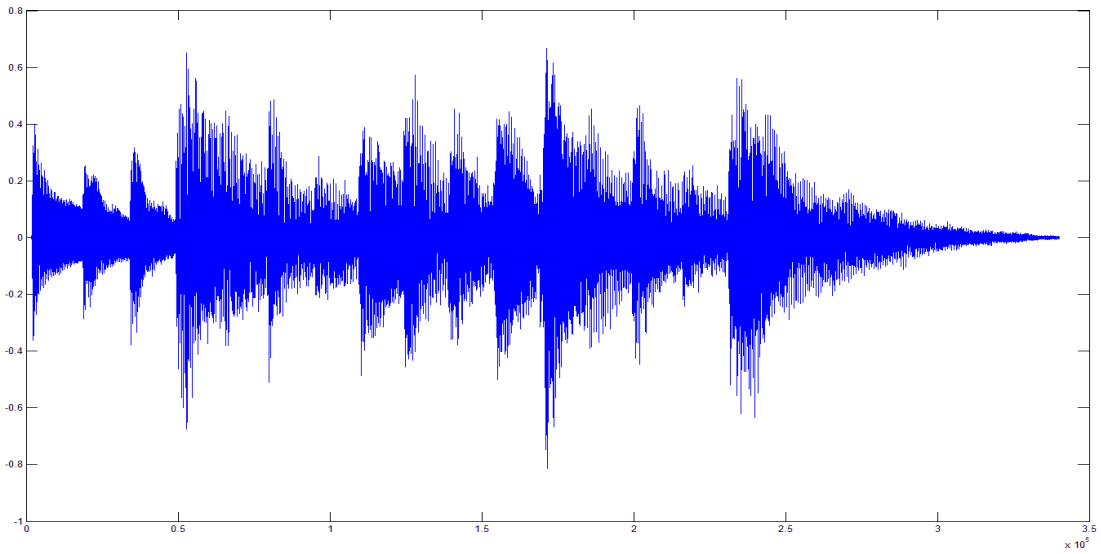

Figure 6. The Original Speech

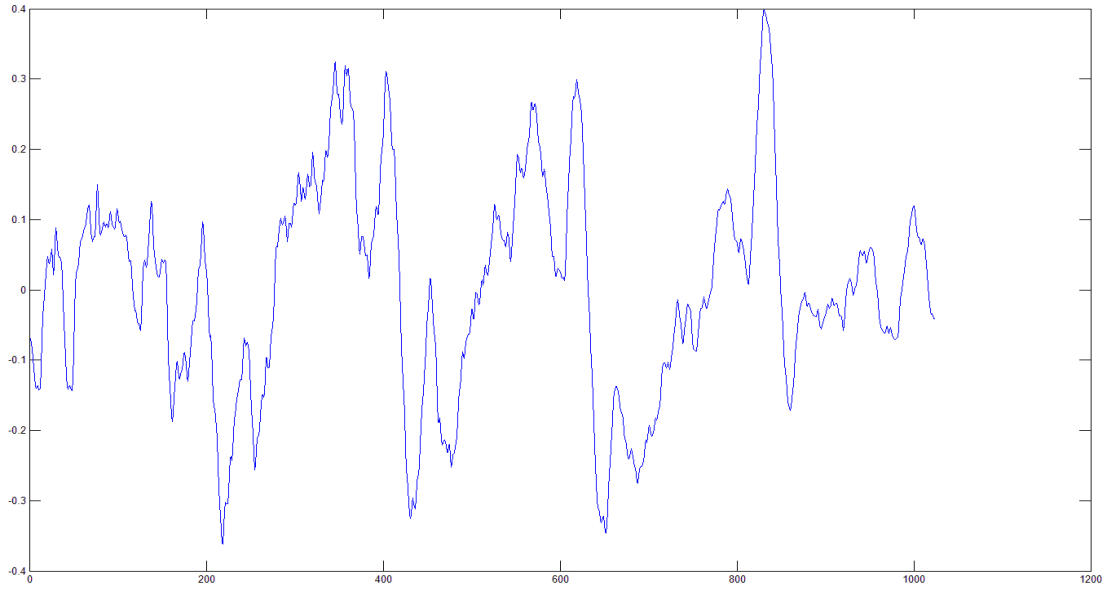

Figure 7. The Test Speech Signal

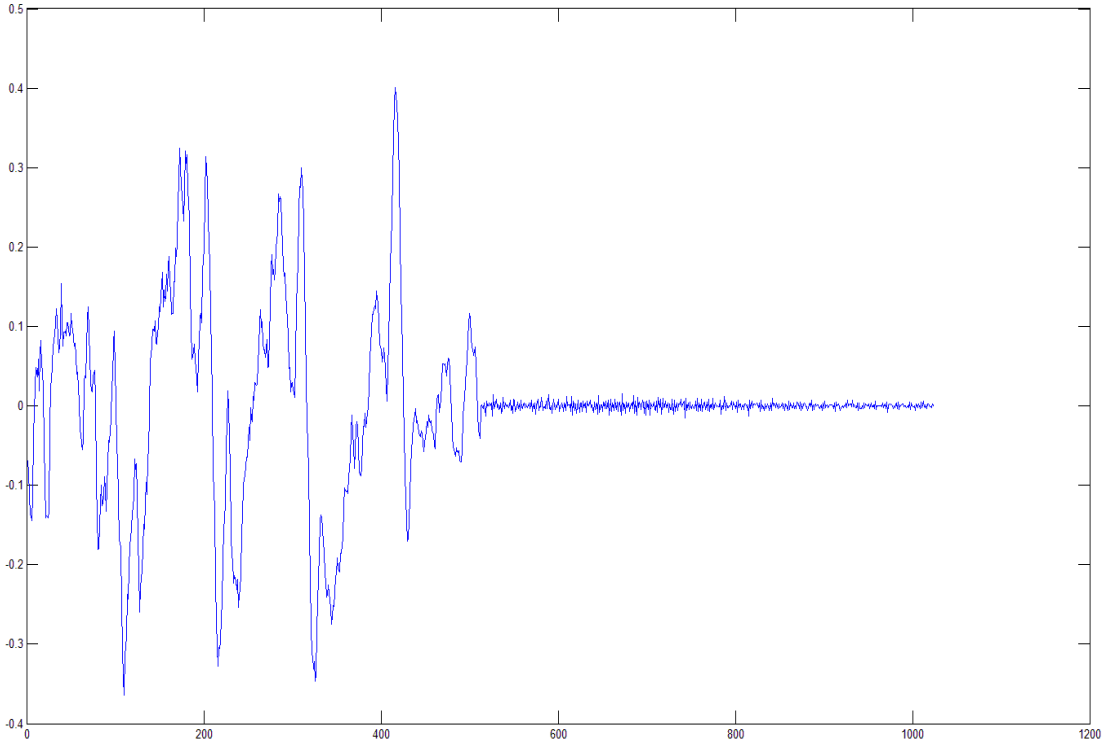

Figure 8. One Level Adaptive Wavelet Decomposition of the Speech Signal

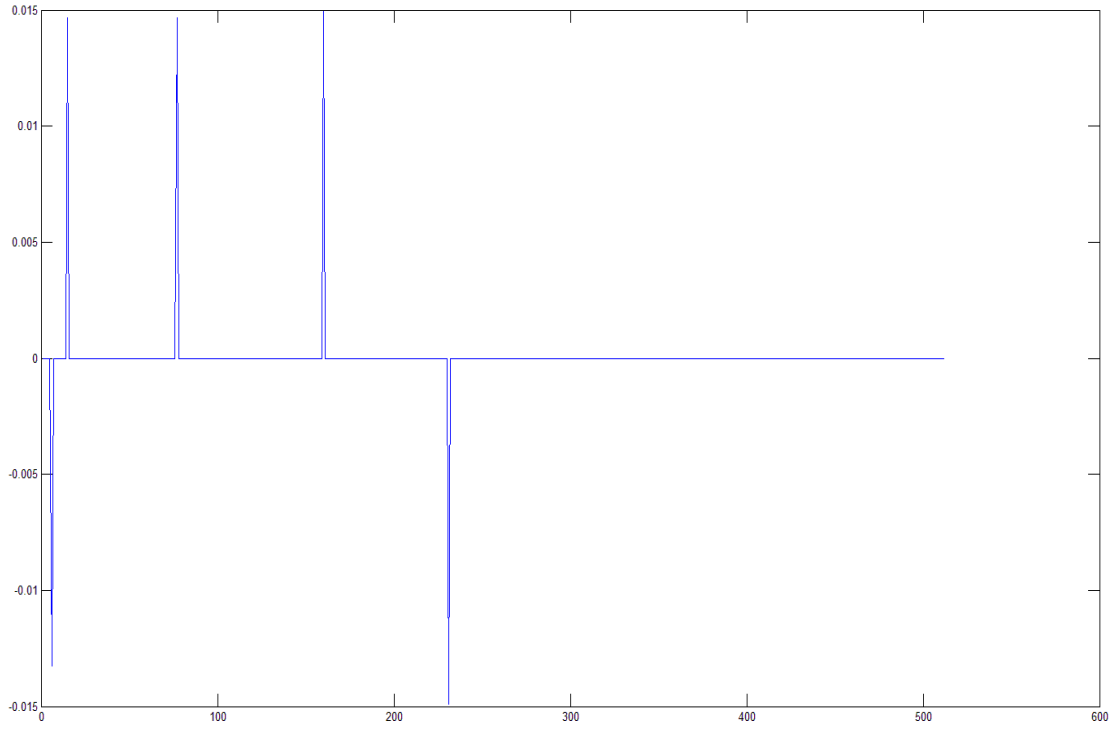

Figure 9. Sparse High Frequency Coefficients

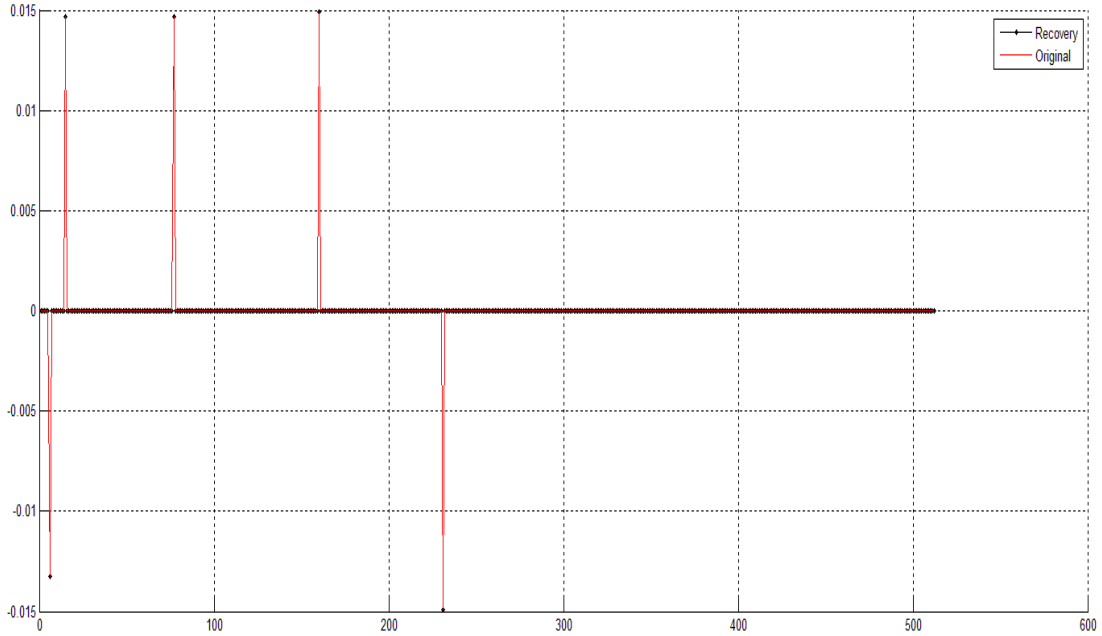

Figure 10. Reconstruction Sparse High Frequecny Coefficients

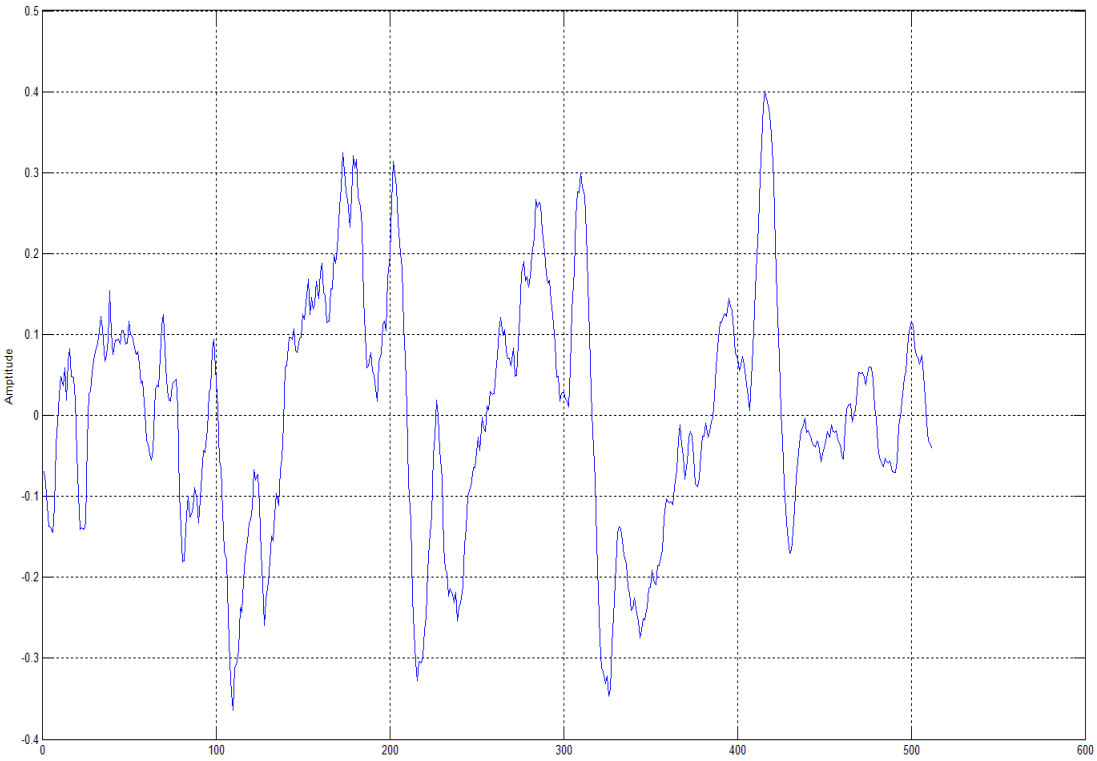

Figure 11. The Low Frequency Coefficients

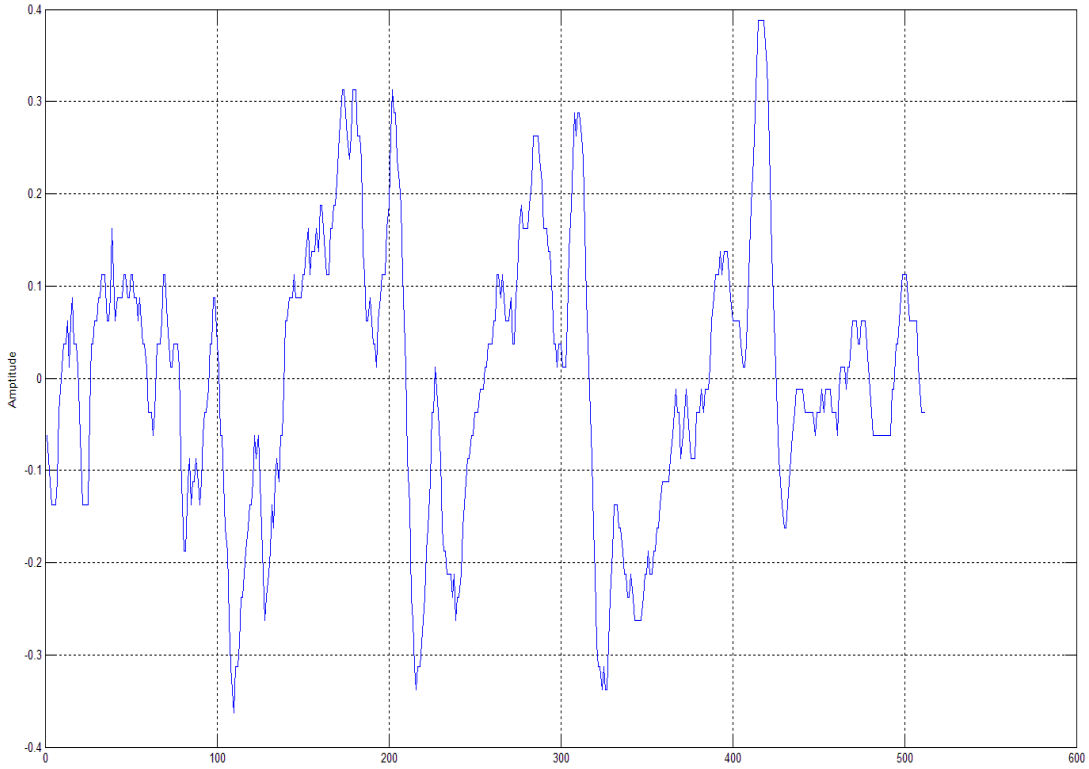

Figure 12. The Quantization Output in PCM for the Low Frequency Coefficients

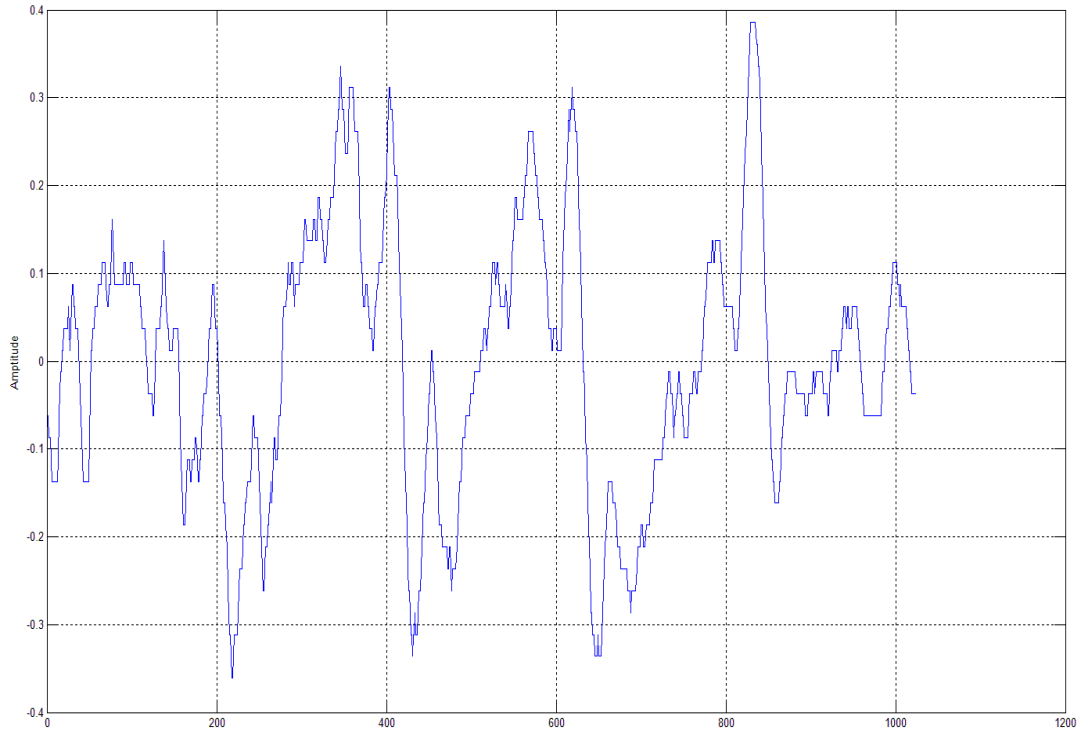

Figure 13. Quantization Output in PCM for All Speech Signal

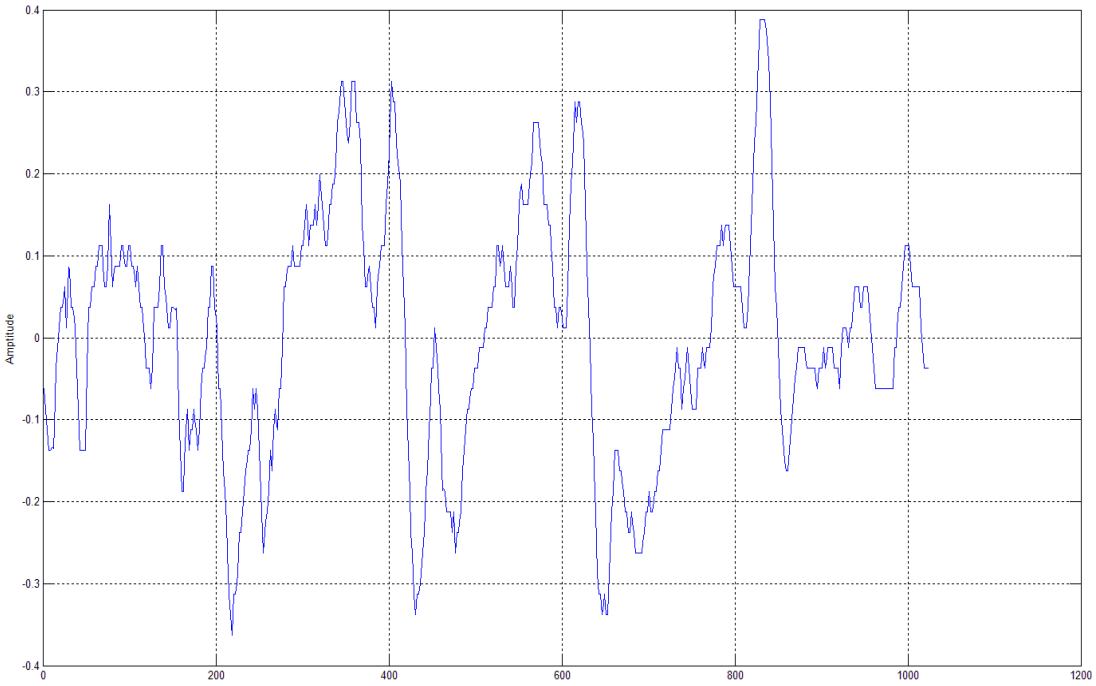

Figure 14. The Joint Codec Output for all Speech Signal

Acknowledgements

This work was supported by Zhejiang Provincial Natural Science Foundation of China (No.Y1110632) and (LY12F01017), Supported by the construct program of the key discipline in Hangzhou.

References

[1] E. Candes and T. Tao, "Decoding by linear programming", IEEE Trans. Inform. Theory, vol. 51, no.12, **(2005)**, pp. 4203-4215.
[2] E. Candes, J. Romberg and T. Tao, "Stable signal recovery from incomplete and inaccurate information", Commun. Pure Appl. Math., vol. 59, no. 9, **(2005)**, pp. 1207-1233.
[3] E. Candes, J. Romberg and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information", IEEE Trans. Inform. Theory, vol. 52, no. 2, **(2006)**, pp. 489-590.
[4] D. Donoho, "Compressed sensing", IEEE Trans. Inform. Theory, vol. 52, no. 4, **(2006)**, pp. 1289-1306.
[5] A. Griffin and P. Tsakalides, "Compressed sensing of audio signals using multiple sensors", Processings of the 16th European signal processing conference **(EUSIPCO'08)**, Lausanne, Switzerland.
[6] T. V. Sreenivas and W. B. Kleijn, "Compressive sensingfor sparsely excited speech signals", in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing **(2009)**, pp. 4125–4128.
[7] D. Giacobello, M. G. Christensen, J. Dahl, S. H. Jensen and M. Moonen, "Sparse linear predictors for speech processing", in Proc. Interspeech, **(2008)**.
[8] G. Peyr´e, "Best Basis Compressed Sensing", IEEE Transaction on Signal Processing, vol. 58, no. 1, **(2010)**, pp. 1-11**.**
[9] E. Candes and D. Donoho, "Curvelets-A surprising effective nonadaptive representation for objects with edges", Curves and Surfaces, Vanderbilt University Press, Nashvielle, **(2000).**
[10] H. Rauhut, K. Schnass and P. Vandergheynst, "Compressed sensing and redundant dictionaries", IEEE Trans. Info. Theory, vol. 54, no. 5, **(2008)**, pp. 2210–2219**.**
[11] M. F. Duarte, M. B. Wakin and R. G. Baraniuk, "Wavelet-domain compressive signal reconstruction using a Hidden Markov Tree model", IEEE International Conference on Acoustics, Speech and Signal Processing **(2008)** March 30-April 4, , pp. 5137-5140, Las Vegas, USA.

[12] C. La and M. Do, "Signal reconstruction using sparse tree representations", Proc. SPIE Wavelets XI **(2005)**, August 30, pp. 273-283.
[13] W.-S. Lu, "Compressed sensing and Sparse Signal processing", University of Victoria, Canada **(2010)**.
[14] W. Sweldens, "The lifting scheme: a construction of second generation wavelets", SIAM J. Math. Anal., vol. 29, no. 2, **(1997)**, pp. 511-546.
[15] R. Claypoole, R. Baraniuk and R. Nowak, "Adaptive wavelet transforms via lifting", Proceedings of the IEEE International Conference on Acoustics, Speech,and Signal Processing **(1998)** May 12-15, pp.1513-1516, Seattle, Washington.
[16] Gao, Guang-chun, Zhao. Sheng-ying, Zhu.Hong-li, Shang. Li-na, "Based on lifting scheme adaptive prediction wavelet transform algorithms", Journal of Circuits and Systems, vol. 15, no. 5, pp. 31-34. **(2010)** (in Chinese).

Authors

Guangchun Gao, he received the PhD degrees in Communication and Information System from the Zhejiang University, in 2004. He is an associate professor of Zhejiang University City College. His research interests include compressed sensing, image processing and coding, and communication signal processing and system design.

Lina Shang, she received M.S.degree in EE of Zhejiang University, in 2006. She is a lecturer of Zhejiang University City College Main research direction is IC design, Wavelet and Compressive Sensing.

Kai Xiong, he received a BSc degree from the Zhejiang University in 1999 and MSc and PhD degrees from the Zhejiang University in 2002 and 2005 respectively, all in Optical Engineering. He currently works for the Zhejiang University City College in Hangzhou,. His research interests include optical imaging, image processing ,and optical measurement.