논문 12-37A-12-13
한국통신학회논문지 '12-12 Vol.37A No.12
http://dx.doi.org/10.7840/kics.2012.37A.12.1122

# Adaptive Compressed Sensing과 Dictionary Learning을 이용한 프레임 기반 음성신호의 복원에 대한 연구

정 성 문[•], 임 동 민[°]

# A Study on the Reconstruction of a Frame Based Speech Signal through Dictionary Learning and Adaptive Compressed Sensing

Seongmoon Jeong[•], Dongmin Lim[°]


## 요 약

압축센싱은 이미지, 음성신호, 레이더 등 많은 분야에 적용되고 있다. 압축센싱은 주로 통계적 특성이 시불변인 신호에 적용되고 있으며, 측정 데이터를 줄여 압축률을 높일수록 복원에러가 증가한다. 이와 같은 문제점들을 해결하기 위해 음성신호를 프레임 단위로 나누어 병렬로 처리하였으며, dictionary learning을 이용하여 프레임들을 sparse하게 만들고, sparse 계수 벡터와 그 복원값의 차를 이용하여 압축센싱 복원행렬을 적응적으로 만든 적응압축센싱을 적용하였다. 이를 통해 통계적 특성이 시변인 신호도 압축센싱을 이용하여 빠르고 정확한 복원이 가능함을 확인할 수 있었다.


**Key Words** : adaptive compressed sensing, dictionary learning, frame based speech signal processing, 적응압축센싱, 프레임 기반 음성신호처리


## ABSTRACT

Compressed sensing has been applied to many fields such as images, speech signals, radars, etc. It has been mainly applied to stationary signals, and reconstruction error could grow as compression ratios are increased by decreasing measurements. To resolve the problem, speech signals are divided into frames and processed in parallel. The frames are made sparse by dictionary learning, and adaptive compressed sensing is applied which designs the compressed sensing reconstruction matrix adaptively by using the difference between the sparse coefficient vector and its reconstruction. Through the proposed method, we could see that fast and accurate reconstruction of non-stationary signals is possible with compressed sensing.


## Ⅰ. Introduction

We use electronic devices such as cell phones, notebook PCs and tablet PCs, and they handle many things happening in our daily life such as buying goods, searching information and sending e-mails. Of those devices, cell phones are one of the necessities, and we spend lots of time with them. In the past, they handled only speech data based on the sampling theory, so data throughput was small. These days, however, it has been gradually increasing due to the convergence of the devices and the increase in the huge amounts of data such as video clips and images, which

• 주저자 : 경상대학교 전자공학과, jony19@kospo.co.kr, 정회원
° 교신저자 : 경상대학교 전자공학과, dmlim@gsnu.ac.kr, 종신회원
논문번호 : KICS2012-08-361, 접수일자 : 2012년 8월 17일, 최종논문접수일자 : 2012년 11월 27일

causes the vast amount of sensing and the consumption of resources. To process and reconstruct the data with a small number of samples, compressed sensing(CS)[1,2] has recently come out. It is a novel paradigm that goes against the common wisdom in data acquisition, and it can be applied to various fields including speech signal processing.

However, it is not suitable for non-stationary signals, and reconstruction error could grow as compression ratios are increased to reduce data processing time. So various methods were used to resolve the problem. In 2009, it was applied to sparsely excited speech signals consisting of successive 40ms frames with a codebook[3], and choosing the sensing matrix adaptively was proposed according to the energy distribution of an original speech signal[4]. In 2011, to speed up the CS process, multicore systems were used with gammatone filterbank and DCT[5].

Here, we propose a different method, adaptive compressed sensing(ACS) with a frame based speech signal, to boost the speed of the CS process and to lower the reconstruction error. First of all, the speech signal is divided into proper frames and processed in parallel. Next, the frames are made sparse by dictionary learning(DL)[6], and ACS is applied. Lastly, to evaluate the performance of the above method, computation time, reconstruction error and Perceptual Evaluation of Speech Quality(PESQ)[11] are compared.

In section Ⅱ, the CS theory is introduced. In section Ⅲ and Ⅳ, parallel processing and design of adaptive compressed sensing reconstruction matrix are introduced. In section Ⅴ, experiments are conducted. Lastly, the conclusion is given.

## Ⅱ. Compressed Sensing Theory

In data acquisition, data are sampled, compressed and transmitted or stored. If they have redundant information, it will cause additional costs and resources. To resolve the problem, CS, a novel paradigm, has been recently released. It starts from the viewpoint that most signals in nature could be represented sparsely, and the sparse coefficients have most information of the signals. Thus, we can reconstruct the signals exactly with high probability by reconstructing only the sparse coefficients from a small amount of linear and non-adaptive measurements.

In brief, CS is a technique to find sparse solutions to underdetermined linear systems. The underdetermined system of linear equations has more unknowns than equations and generally has an infinite number of solutions. It is required to find a sparse and unique solution of the following the equation

$$y = \Phi x = \Phi \Psi \alpha = \Theta \alpha \tag{1}$$

where an $N \times 1$ signal vector $x$ is $K$-sparse (explained in section 2.1), $\Phi$ is an $M \times N$ measurement matrix which is in charge of sensing and compression, $y$ is an $M \times 1$ measurement vector, $\Psi$ is an $N \times N$ representation matrix which makes the signal $x$ sparse, $\Theta$ is an $M \times N$ compressed sensing reconstruction matrix, and $\alpha$ is an $N \times 1$ coefficient vector ( $K \ll M < N$ ).

### 2.1. Sparse representation

Signals can be represented concisely in the $\Psi$ domain, for example, DFT of speech signals. They can be expressed as

$$x = \sum_{i=1}^{n} \alpha_i \psi_i \quad \text{or} \quad x = \Psi \alpha, \ \alpha_i = < x, \psi_i > \tag{2}$$

where $\alpha_i$ is the coefficients of $\alpha$, and $\psi_i$ is the columns of $\Psi$. If only $K$ of $\alpha_i$ coefficients are nonzero, the signal $x$ is $K$-sparse, and if the coefficients consist of many small elements and a few large elements, the signal $x$ is called compressible. Through the sparse representation, we can obtain the efficiency of signal acquisition and the reduction of the required resource for storage or transmission.

### 2.2. RIP and incoherence

If the location of $K$ nonzero coefficients is known, the reconstruction of the signal $x$ can be done with the $M \times N$ measurement matrix $\Phi$ ($M \ll N$). Otherwise, it is hard to reconstruct $x$. However, if $\Phi$ satisfies Restricted Isometry Property(RIP)[1,2], accurate reconstruction is possible. It is defined as

$$(1-\delta_k)\,\|x\|_2^2 \le \|\Phi x\|_2^2 \le (1+\delta_k)\,\|x\|_2^2 \tag{3}$$

where $\delta_k$ is an isometry constant, and $0 < \delta_k < 1$. RIP means that $\Phi$ projects uniform energy to each element of $x$. An important thing is that $\Phi$ should project energy uniformly for an arbitrary $x$. Furthermore, there is another factor for accurate reconstruction called incoherence[2]. It determines measurement $M$ below

$$M \ge C\mu^2(\Phi,\Psi)K\log N \tag{4}$$

where $C$ is a constant, $K$ is the sparsity level, and the coherence between $\Phi$ and $\Psi$, $\mu(\Phi,\Psi)$, is defined as

$$\mu(\Phi,\Psi) = \sqrt{n} \cdot \max_{1 \le k,j \le n} |<\varphi_k,\psi_j>| \tag{5}$$

where $\mu(\Phi,\Psi) \in [1, \sqrt{n}]$, and if $\Phi$ and $\Psi$ include correlated elements, the coherence is large. Generally, a random matrix is chosen as $\Phi$, which satisfies both RIP and incoherence.

### 2.3. Reconstruction

If RIP holds or $\Phi$ and $\Psi$ are incoherent to each other, the signal can be reconstructed by $L1$ norm minimization as

$$\hat{\alpha} = \arg\min \|\alpha\|_1 \quad s.t. \quad y = \Theta\alpha \tag{6}$$

where if $M \ge C\mu^2(\Phi,\Psi)K\log N$, exact reconstruction is done with overwhelming probability[2]. $L1$ norm minimization is a convex optimization problem, which can be substituted with a linear programming method called Basis Pursuit(BP)[8]. Nowadays, in addition, an iterative greedy algorithm, Orthogonal Matching Pursuit(OMP)[7], is used frequently due to its fast and accurate reconstruction.

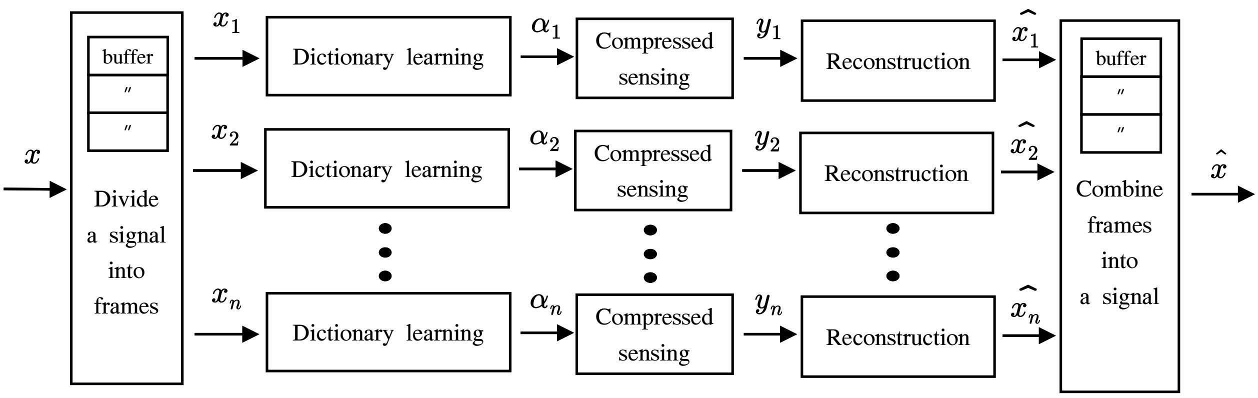


Fig. 1. Implementation of the framed speech signal experiment
$x$ : signal, $x_1$, $x_2$, $x_n$ : 1st, 2nd, *n*th frame, $\alpha_1$, $\alpha_2$, $\alpha_n$ : 1st, 2nd, *n*th sparse coefficients
$y_1$, $y_2$, $y_n$ : 1st, 2nd, *n*th observed frame, $\hat{x}_1$, $\hat{x}_2$, $\hat{x}_n$ : 1st, 2nd, *n*th reconstructed frame, $\hat{x}$ : reconstructed signal

## Ⅲ. Parallel Processing

To speed up the CS process, it can be processed in parallel as in Fig. 1. Given a signal $x$, it is divided into $n$ frames where $x_k$ is the $k$th frame. Each frame is made sparse by a DL method. Then sparse coefficient vectors go through the compression process. Lastly, all sparse coefficient vectors are reconstructed into frames and they are combined into the signal $\hat{x}$. In this processing, we use buffers to store signal frames, which reduce the number of the frame compression and reconstruction process. For example, a signal is divided into 10 frames and five compression and reconstruction processes are processed two times.

## Ⅳ. Design of Adaptive Compressed Sensing Reconstruction Matrix

In CS, the measurement matrix affects compression and the result of reconstruction. As we choose proper one which satisfies RIP, good reconstruction results could come out. Generally, the random matrix is used, but it is not suitable for every signal. So we propose the adaptive compressed sensing reconstruction matrix, which modifies the matrix by adding the difference between the sparse coefficient vector and its reconstruction. This process compensates for the error occurring by the compression ratio, but it needs the reconstructed signal. It is different from the method of the reference [4] that explains the design of the measurement matrix adaptively according to the energy distribution of an original speech signal before transmitting the compressed signal. In short, our method enhances performance by designing the compressed sensing reconstruction matrix adaptively through compensation for errors resulting from compression and reconstruction. The proposed method shown in Fig. 2 is defined as follows.

i ) Initialize : $i=-1$, $\widehat{\alpha}'^{-1}=\alpha$, $\Theta'^{-2}=\Theta=\Phi D$, $y'^{-2}=y=\Theta\alpha$ where $i$ is the iteration number, $\alpha$ is an $N\times 1$ sparse coefficient vector, and $\widehat{\alpha}'^{i}$ is the reconstructed one with a newly designed compressed sensing reconstruction matrix $\Theta'^{i}$ in iteration $i$. An $M\times N$ compressed sensing reconstruction matrix is denoted by $\Theta=[\theta_1,\theta_2,\dots,\theta_N]$, where $\theta_k$ is the $k$th column vector $(M\ll N)$. $\Phi$ is an $M\times N$ measurement matrix, $D$ is an $N\times N$ dictionary matrix, and $y'^{i}$ is the $M\times 1$ measurement vector by $\Theta'^{i}$ in iteration $i$.

ii ) Repeat then $i=i+1$ (If $i=-1$, go to step iii).

iii ) Obtain the sparse coefficients

$$\widehat{\alpha}'^{i}=\Theta'^{i-1\dagger}y'^{i-1} \tag{7}$$

where $\Theta'^{i\dagger}=\Theta'^{iT}(\Theta'^{i}\Theta'^{iT})^{-1}$.

iv ) Obtain the error between the sparse coefficient vector and its reconstruction

$$e^{i}=\alpha-\widehat{\alpha}'^{i} \tag{8}$$

v ) Add the $k$th element of the error vector to the $k$th column of the compressed sensing reconstruction matrix and repeat it until $k=N$

$$\theta'^{i}_{k}=\theta'^{i-1}_{k}+e^{i}_{k}\ ,\ k=1,\dots,N \tag{9}$$

where $e^{i}_{k}$ is the $k$th element of the error vector in iteration $i$, and $\theta'^{i}_{k}$ is the $k$th column vector of $\Theta'_{k}$ in iteration $i$.

vi) Design the new compressed sensing reconstruction matrix by combining the column vectors in ascending order of indices

$$\Theta'^{i}=[\,\theta'^{i}_{1}\,\theta'^{i}_{2},\dots,\theta'^{i}_{N}]. \tag{10}$$

vii) Obtain the measurements

$$y'^{i}=\Theta'^{i}\alpha. \tag{11}$$

viii) Repeat from step ii ) until $\|\alpha-\widehat{\alpha}'^{i}\|_2\le\rho$, where $\rho$ is an error value for termination. If $i=-1$, go to step ii ).

ix) Reconstruct the signal

$$\begin{aligned}\widehat{\alpha}&=\Theta'^{i\dagger}y'^{i}\\ \widehat{x}&=D\widehat{\alpha}\end{aligned} \tag{12}$$

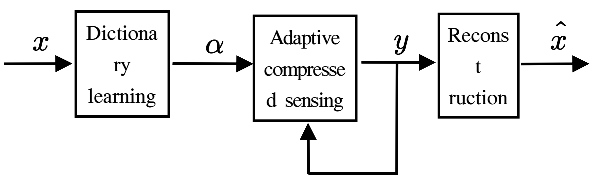


Fig. 2. Implementation of the proposed algorithm
$x$ : signal, $\alpha$ : sparse coefficients
$y$ : measurements, $\hat{x}$ : reconstructed signal

## V. Experiments

Files "47242_alphahog_annabel-lee-2.mp3" and "34210_ acclivity_i-am-female.mp3" (6,133,248 and 6,220,800 samples respectively with sampling rate of 44.1kHz) were downloaded from "www.freesound.org" for experiments. The former is a male speech signal and the latter is a female speech signal. They were transformed to wave files and downsampled to 16kHz. Their lengths were reduced properly. For the sparse representation, greedy adaptive dictionary(GAD)[6], [12] was used as a dictionary learning method. It provides good performance in sparse representation, computation time, and reconstruction error in comparison with *K*-SVD[9] and principal component analysis(PCA) for speech signals. For reconstruction, OMP[10] was used, which provides better performance in computation time and reconstruction error in comparison with BP.

### 5.1. Parallel compressed sensing

As shown in Fig. 1, we divided the female speech signal into frames and compared the framed speech signal with unframed one in terms of computation time, reconstruction error and PESQ by changing compression ratios from 0 to 1 and averaged over 100 times under the conditions of Table 1.

Table 1. Conditions for the experiments

<table>
<tr><th colspan="2">Framed speech signal</th><th colspan="2">Unframed speech signal</th></tr>
<tr><td rowspan="5">Frame (duration time per frame, ms)</td><td>3 (44,000)</td><td rowspan="5">Frame (duration time per frame, ms)</td><td rowspan="5">1 (13,2000)</td></tr>
<tr><td>12 (11,000)</td></tr>
<tr><td>27 (4,888.89)</td></tr>
<tr><td>48 (2,750)</td></tr>
<tr><td>108 (1,222.22)</td></tr>
</table>

As the number of frames is increased, the reconstruction error grows as shown in Fig. 3. Reconstruction error $\epsilon$[6] is defined as

$$\epsilon = \| x(t) - \hat{x}(t) \|_2 \tag{13}$$

The more frames are divided, the more reconstruction is required. During this process, OMP causes increase in errors. In result, a trade-off between the reconstruction error and the number of frames is required. If fast reconstruction is needed regardless of the error, it is desirable to increase the number of frames. If accurate reconstruction is required, on the contrary, the number of frames should be reduced. Therefore, it is necessary to choose the proper number of frames for fast and accurate reconstruction.

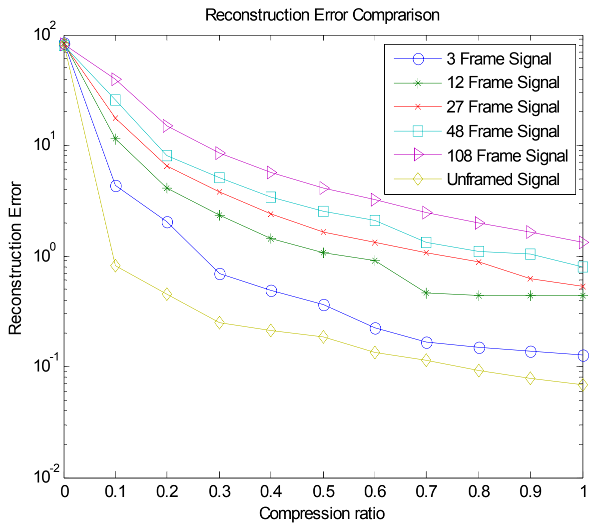


Fig. 3. Reconstruction error of frame based signals and an unframed signal

As expected, the framed one is faster than the unframed one. As shown in Fig. 4, the more frames are, the faster computation time is. We might think that if the speech signal is divided into 3 frame and processed in parallel, it will be three times faster. However, the result shows it is almost twelve times faster in average computation time.

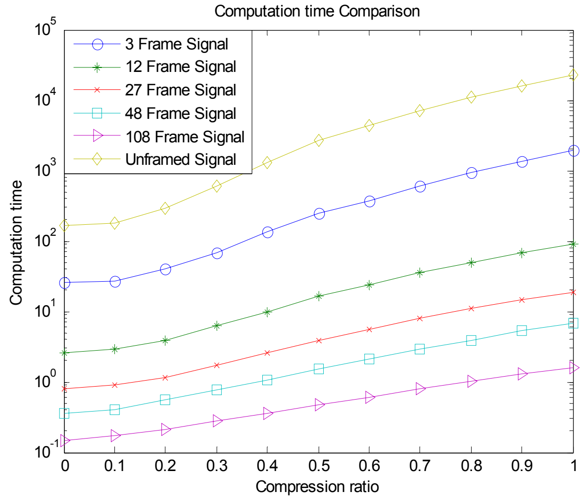


Fig. 4. Computation time of framed signals and an unframed signal

PESQ is an objective method to compare an original signal with an attenuated signal coming out from a system. Its value ranges from -0.5 to 4.5 and normally has the value from 1 to 4.5. As shown in Fig. 5, PESQ decreases when the number of frames is increased. When the compression ratio is 0.3, PESQ is 3.97 for 3 frames, 3.55 for 12 frames, and 3.24 for 27 frames.

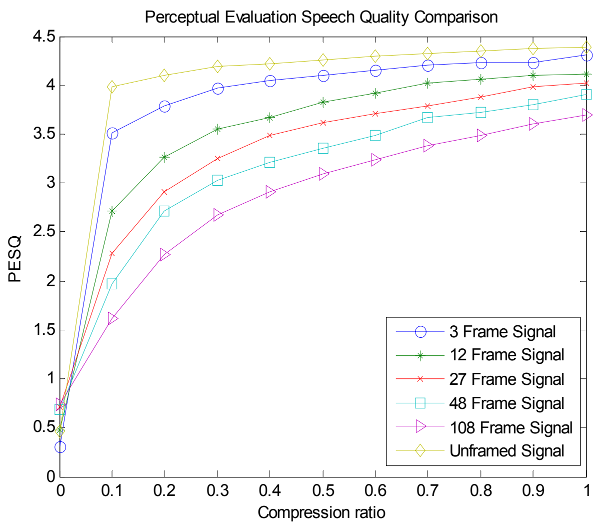


Fig. 5. PESQ of framed signals and an unframed signal

As a result of the above experiments, reconstruction error and PESQ get worse as the number of frames grows, but computation time is opposite to those. Compression ratios for accurate reconstruction were different for speech signals due to their different sparsity. Generally, it is possible to reconstruct speech signals accurately with the compression ratios between 0.2 and 0.4. Additionally, we compared the male speech signal with the female one, but conditions such as signal waveform, sparsity, length and so on were different so it was not easy to compare them.

### 5.2. Adaptive compressed sensing

We applied ACS to the frame based speech signal. The female speech signal was divided into 5,292 frames, and the duration time of each frame was 25ms. The duration time of 20～40ms is proper for CS of non-stationary speech signals because non-stationarity can be overcome with short frames. The iteration was set to 0. Reconstruction error, computation time and PESQ were measured by changing the compression ratio from 0 to 1 and averaged over 100 times. In the experiments, four ways were compared as follows:

1. Non-adaptive compressed sensing(NACS) with the framed speech signal
2. Non-adaptive compressed sensing(NACS) with the unframed speech signal
3. Adaptive compressed sensing(ACS) with the framed speech signal
4. Adaptive compressed sensing(ACS) with the unframed speech signal

As shown in Fig. 6, the reconstruction performance of ACS is better than that of NACS for framed speech signals, but for the unframed ones with long duration, the ACS performance is worse than that of NACS in some compression ratios. Additionally, we conducted several experiments and discovered that ACS performance is better for the frames below 250,000 samples (about 15.6 seconds).

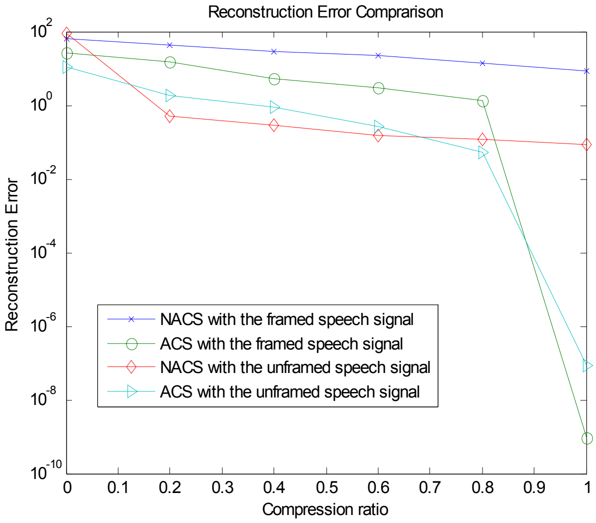


Fig. 6. Reconstruction error of ACS and NACS

Computation time of ACS and NACS is compared in Fig. 7. Average computation time per frame is 0.002 seconds for NACS with the framed speech signal, 0.006 seconds for ACS with the framed speech signal, 756.28 seconds for NACS with the unframed speech signal and 1,496.20 seconds for ACS with the unframed speech signal. In terms of PESQ, the performance by ACS is improved greatly and better than that by NACS as shown in Fig. 8.

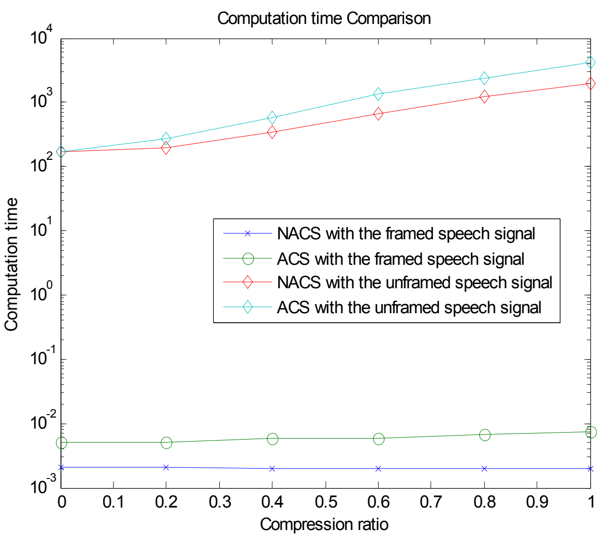


Fig. 7. Computation time of ACS and NACS

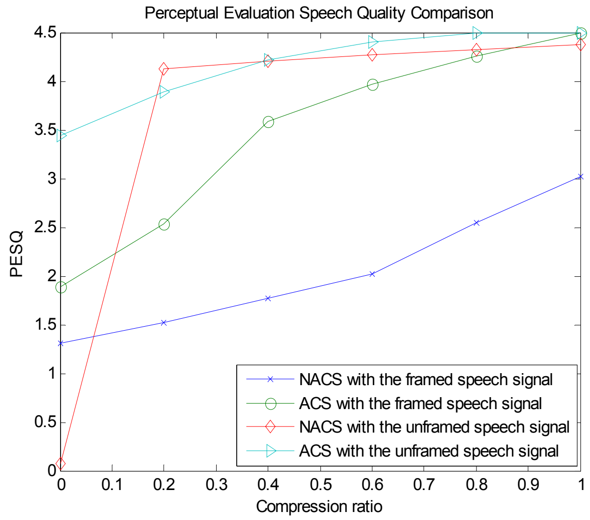


Fig. 8. PESQ of ACS and NACS

Furthermore, as shown in Fig. 9, ACS with unframed speech signal was iterated once to see the effect of iteration and it was found that the reconstruction error decreased dramatically. In result, reconstruction error and PESQ are improved, and computation time is increased by ACS. Here, the reconstruction error, PESQ and computation time could be enhanced at once by increasing the number of frames and the number of iterations of the ACS process properly. As further work, we plan to do research on ACS in noisy environment.

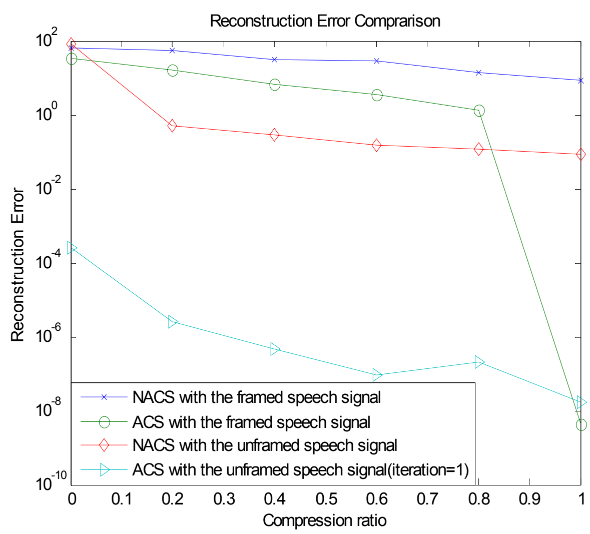


Fig. 9. Reconstruction error of ACS and NACS with iteration=1

## Ⅵ. Conclusion

Compressed sensing is a novel paradigm that does sampling and compression at once and reconstructs signals with fewer samples than the sampling theory requires. These days, CS holds the spotlight, and a lot of research is conducted actively. However, CS is not suitable for non-stationary signals, and reconstruction error grows as the compression ratios are increased to reduce data processing time.

In this paper, we made an attempt to resolve the problem. First, speech signals were divided into frames and processed in parallel. Next, the frames were made sparse by DL, and ACS was applied. As a result, when the number of frames was increased, the reconstruction error increased, but the computation time and PESQ were in inverse proportion to the number of frames. As ACS was applied, the performance in the reconstruction error and PESQ was enhanced dramatically. Furthermore, the performance improves with more iterations of the ACS process.

In conclusion, fast and accurate reconstruction is possible for non-stationary speech signals by increasing the number of frames, and applying ACS although there is a trade-off between reconstruction error and computation time. With those merits of low reconstruction error with high compression ratio and fast processing time, the proposed method could be applied to DSP for real-time processing and transmission for big data and data storage techniques. Hence, it could cause the changes of electronic devices in terms of smaller size and lower power consumption.

## Appendix

Let $\Theta$ be an $M\times N$ matrix, $\alpha$ be an $N\times 1$ vector, and $y$ be an $M\times 1$ vector.

$$\Theta=\begin{bmatrix}\theta_{11} & \theta_{12} & \cdots & \theta_{1N}\\ \theta_{21} & \theta_{22} & \cdots & \theta_{2N}\\ \vdots & \vdots & \ddots & \vdots\\ \theta_{M1} & \theta_{M2} & \cdots & \theta_{MN}\end{bmatrix},\ \alpha=\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_N\end{bmatrix},\ y=\Theta\alpha=\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_M\end{bmatrix} \tag{14}$$

Given a $1\times 5$ matrix $\Theta$ (CS focuses on high compression ratios) and a $5\times 1$ vector $\alpha$, we can reconstruct $\alpha$ with the pseudo inverse, then we obtain the results as follows.

$$\Theta=\begin{bmatrix}\theta_{11} & \theta_{12} & \theta_{13} & \theta_{14} & \theta_{15}\end{bmatrix} \tag{15}$$

$$y=\Theta\alpha=\begin{bmatrix}\theta_{11}\alpha_1+\theta_{12}\alpha_2+\theta_{13}\alpha_3+\theta_{14}\alpha_4+\theta_{15}\alpha_5\end{bmatrix} \tag{16}$$

$$\widehat{\alpha}=\Theta^{\dagger}y=\begin{bmatrix}\widehat{\alpha_1}\\ \widehat{\alpha_2}\\ \widehat{\alpha_3}\\ \widehat{\alpha_4}\\ \widehat{\alpha_5}\end{bmatrix} \tag{17}$$

$$\widehat{\alpha_1}=\frac{\theta_{11}\sum_{i=1}^{5}\theta_{1i}\alpha_i}{\sum_{i=1}^{5}\theta_{1i}^2},\quad \widehat{\alpha_2}=\frac{\theta_{12}\sum_{i=1}^{5}\theta_{1i}\alpha_i}{\sum_{i=1}^{5}\theta_{1i}^2},$$
$$\widehat{\alpha_3}=\frac{\theta_{13}\sum_{i=1}^{5}\theta_{1i}\alpha_i}{\sum_{i=1}^{5}\theta_{1i}^2},\quad \widehat{\alpha_4}=\frac{\theta_{14}\sum_{i=1}^{5}\theta_{1i}\alpha_i}{\sum_{i=1}^{5}\theta_{1i}^2},$$
$$\widehat{\alpha_5}=\frac{\theta_{15}\sum_{i=1}^{5}\theta_{1i}\alpha_i}{\sum_{i=1}^{5}\theta_{1i}^2} \tag{18}$$

where $\Theta^{\dagger}=\Theta^T(\Theta\Theta^T)^{-1}$. Through the equations (8)~(10), we can get the error $e$ and $\Theta'$. By using $\Theta'$, the vector $\alpha$ is reconstructed as follows

$$e=\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix},$$

$$\Theta' = \left[\theta_{11}+e_1\ \theta_{12}+e_2\ \theta_{13}+e_3\ \theta_{14}+e_4\ \theta_{15}+e_5\right] \quad (19)$$

$$\widehat{\alpha}' = \Theta'^{\dagger} y' = \begin{bmatrix} \widehat{\alpha_1}' \\ \widehat{\alpha_2}' \\ \widehat{\alpha_3}' \\ \widehat{\alpha_4}' \\ \widehat{\alpha_5}' \end{bmatrix}, \quad y' = \Theta' \alpha \quad (20)$$

$$\widehat{\alpha_1}' = \frac{(e_1+\theta_{11})\left\{\sum_{i=1}^{5}(\alpha_i(e_i+\theta_{1i}))\right\}}{\sum_{i=1}^{5}(e_i+\theta_{1i})^2},$$

$$\widehat{\alpha_2}' = \frac{(e_2+\theta_{12})\left\{\sum_{i=1}^{5}(\alpha_i(e_i+\theta_{1i}))\right\}}{\sum_{i=1}^{5}(e_i+\theta_{1i})^2},$$

$$\widehat{\alpha_3}' = \frac{(e_3+\theta_{13})\left\{\sum_{i=1}^{5}(\alpha_i(e_i+\theta_{1i}))\right\}}{\sum_{i=1}^{5}(e_i+\theta_{1i})^2},$$

$$\widehat{\alpha_4}' = \frac{(e_4+\theta_{14})\left\{\sum_{i=1}^{5}(\alpha_i(e_i+\theta_{1i}))\right\}}{\sum_{i=1}^{5}(e_i+\theta_{1i})^2},$$

$$\widehat{\alpha_5}' = \frac{(e_5+\theta_{15})\left\{\sum_{i=1}^{5}(\alpha_i(e_i+\theta_{1i}))\right\}}{\sum_{i=1}^{5}(e_i+\theta_{1i})^2} \quad (21)$$

Let's take a look at the equation (18) and (21) in detail. The difference is that the *i*-th element of $e$ is added to the *i*-th column of $\Theta$. It means that $e$ determines $\widehat{\alpha}'$. For the sake of a simple explanation, let $\alpha_1$ be positive, and $\alpha_2$, $\alpha_3$, $\alpha_4$ and $\alpha_5$ are negative as in Fig. 10. With the high compression ratio of 0.2, the element values of $\widehat{\alpha}$ are close to zero (See Fig. 10(b)). Then the element values of $e$ approach those of $\alpha$, and the element signs of $e$ are the same as those of $\alpha$.

In the equation (21), we can consider three cases about the relations between $e_i$ and $\theta_{1i}$ as follows.

1. All values of $e_i$ are bigger than those of $\theta_{1i}$.
2. Some values of $e_i$ are bigger than those of $\theta_{1i}$, and the others are smaller.
3. All values of $e_i$ are smaller than those of $\theta_{1i}$.

As all mathematical formulas are similar in the equation (21), we consider the case of $\widehat{\alpha_1}'$, and $\theta_{1i}$ are small enough as we use a $1/\sqrt{M}\times$ random matrix($M\times N$) as the measurement matrix $\Phi$.

First of all, let us see the first case. All signs of $\theta_{1i}+e_i$ follows those of $e_i$. Each term $a_i(e_i+\theta_{1i})$ in the braces of the numerator becomes positive, and the values of $a_i(e_i+\theta_{1i})$ are almost close to those of $(e_i+\theta_{1i})^2$ in the denominator so the first term $(e_1+\theta_{11})$ in the numerator determines the decreases or increases in $\widehat{\alpha_1}'$. In the second case, as some values which are bigger than those of $\theta_{1i}$ in $e_i$ follow the first case, we only consider the others which are smaller than those of $\theta_{1i}$ in $e_i$.

Under the condition that the values of $e_i$ are almost the same as those of $\alpha_i$ in high compression ratios, and $\theta_{1i}$ are small enough, the absolute values of terms $a_i(e_i+\theta_{1i})$ in the braces of the numerator are close to those of $(e_i+\theta_{1i})^2$, and all values are small enough to be neglected. Thus, $\widehat{a_1}' \risingdotseq e_1+\theta_{11}$. The same principle of the above case can be applied to the last case. In result, ACS compensates for the error between the original signal and its reconstruction with high probability.

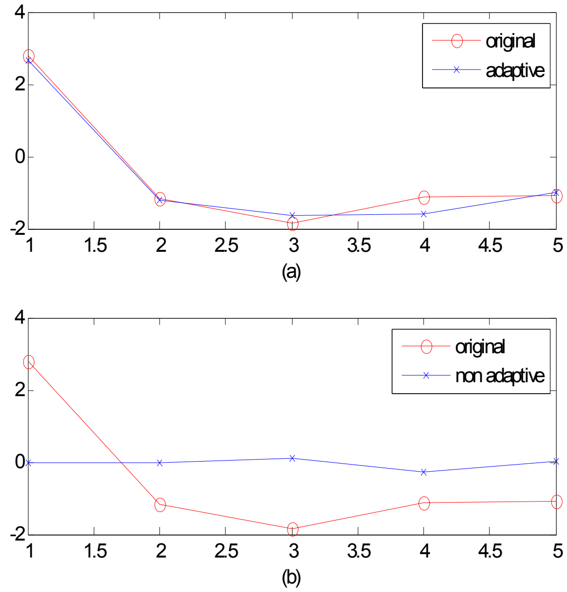


Fig. 10. Implementation of the above mathematical explanation, $\alpha$=5×1 vector, $\Theta$=5×5 random matrix, $M/N$=1/5, iteration=0
(a) Comparison of an original signal $\alpha$ and its reconstruction by ACS
(b) Comparison of an original signal $\alpha$ and its reconstruction by NACS

Moreover, when the above process is iterated, as it compensates for the error repeatedly, more accurate reconstruction could be obtained. According to extra experiments, two or three iterations were proper for accurate reconstruction. Furthermore, ACS can be used in dense signals as in Fig. 12, and unless $\alpha$ exceeds the particular dimension, good results could come out even in high compression ratios without iterations. Fig. 10～12 show the implementation of the above mathematical explanation, where (a) parts of the figures are the comparisons between the original signal $\alpha$(red line) and its reconstruction by ACS(blue line), and (b) parts of the figures are the comparisons between an original signal $\alpha$ (red line) and its reconstruction by NACS(blue line).

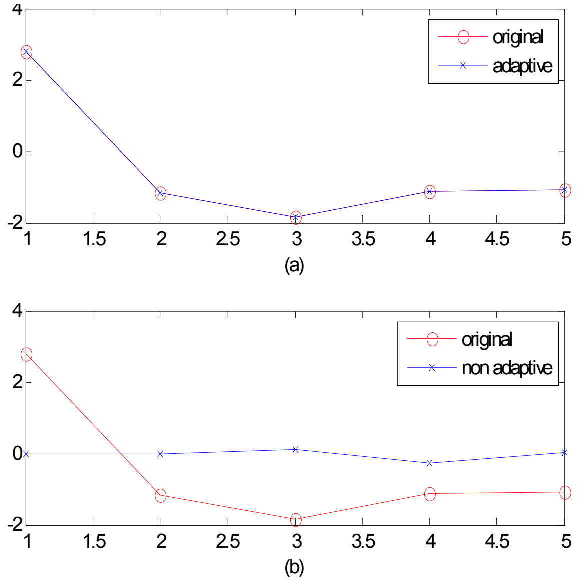


Fig. 11. Implementation of the above mathematical explanation, $\alpha$=5×1 vector, $\Theta$=5×5 random matrix, $M/N$=1/5, iteration=2
(a) Comparison of an original signal $\alpha$ and its reconstruction by ACS
(b) Comparison of an original signal $\alpha$ and its reconstruction by NACS

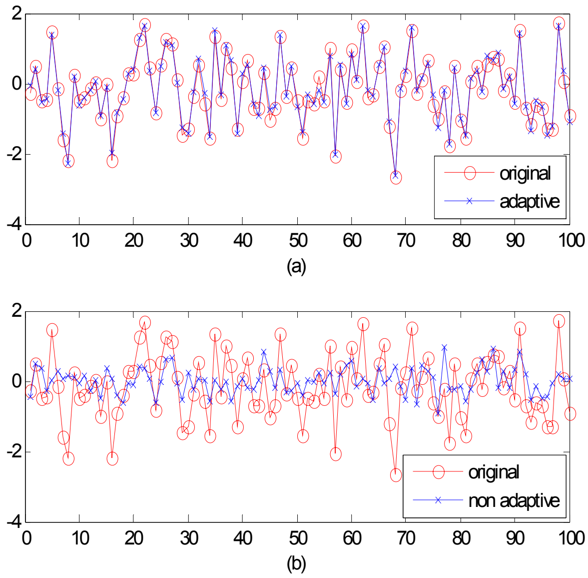


Fig. 12. Implementation of the above mathematical explanation, $\alpha$=100×1 vector, $\Theta$=100×100 random matrix, $M/N$=1/10, iteration=0
(a) Comparison of an original signal $\alpha$ and its reconstruction by ACS
(b) Comparison of an original signal $\alpha$ and its reconstruction by NACS

## References

[1] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Proc. Mag.*, vol. 24, no. 4, pp. 118-121, Jul. 2007.

[2] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Proc. Mag.*, vol. 25, no. 2, pp. 21-30, Mar. 2008.

[3] T. V. Sreenivas and W. B. Kleijn, "Compressive sensing for sparsely excited speech signals," in *Proc. ICASSP*, pp. 4125-4128, Apr. 2009.

[4] T. Xu, Z. Yang and X. Shao, "Adaptive compressed sensing of speech signal based on data-driven dictionary," in *Proc. APCC 2009*, pp. 257-260, Sep. 2009.

[5] T. S. Gunawan, O. O. Khalifa, A. A. Shafie and E. Ambikairajah, "Speech compression using compressive sensing on a multicore system," in *Proc. ICOM 2011*, May 2011.

[6] M. G. Jafari and M. D. Plumbley, "Speech denoising based on a greedy adaptive dictionary algorithm," in *Proc. European Signal Proc. Conf.*, pp. 1423-1426, Aug. 2009.

[7] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[8] S. S. Chen, D. L. Donoho and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129-159, 2001.

[9] M. Ahron, M. Elad and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representations," *IEEE Trans. on Signal Proc.*, vol. 54, no. 11, pp. 4311-4322, Nov. 2006.

[10] K-SVD and OMP Matlab toolbox, from http://www.cs. technion.ac.il/~elad/software/.

[11] ITU-T Study Group 12, "Perceptual evaluation of speech quality(PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," *ITU-T Recommendation* pp. 862, Feb. 2001.

[12] M. G. Jafari and M. D. Plumbley, "Fast dictionary learning for sparse representations of speech signals," *IEEE signal proc.*, vol. 5, no. 5, pp. 1027-1030, Sep. 2011.

[13] S. Kunis and H. Rauhut, "Random sampling of sparse trigonometric polynomials, II. Orthogonal matching pursuit versus basis pursuit," *Foundations of Computational Mathematics*, vol. 8, no. 6, pp. 737-763, Dec. 2008.

**정 성 문** (Seongmoon Jeong)

2007년 2월 동아대학교 전자공학과 공학사

2012년 8월 경상대학교 전자공학과 공학석사

2007년 3월~현재 한국남부발전

<관심분야> 압축센싱, 음성신호처리

**임 동 민** (Dongmin Lim)

1986년 2월 서울대학교 제어계측공학과 공학사

1988년 2월 한국과학기술원 전기 및 전자공학과 공학석사

1992년 8월 한국과학기술원 전기 및 전자공학과 공학박사

2010년 7월~2011년 6월 미국 오하이오주립대 전기 및 컴퓨터공학과 객원연구원

1992년 9월~현재 경상대학교 전자공학과 교수, 경상대학교 부속 공학연구원 연구원

<관심분야> 통신신호처리, RF & DSP 시스템 설계, 이동통신