REDUCTION OF AUTOREGRESSIVE NOISE WITH SHIFT-INVARIANT WAVELET-PACKETS

Nathaniel A. Whitmal and Janet C. Rutledge *

Dept. of EECS
Northwestern University
2145 Sheridan Road
Evanston, IL 60208-3118

Jonathan Cohen †

Dept. of Mathematical Sciences
DePaul University
2219 N. Kenmore
Chicago, IL 60614-3504

ABSTRACT

In this paper, we present a new wavelet-based method for reducing additive autoregressive noise. The method uses a shift-invariant wavelet-packet transform to facilitate a linear transformation of wavelet-packet basis vectors. The transformed basis vectors are shown to be better suited than the original basis vectors for use in conventional wavelet-based denoising algorithms which use MDL or thresholding approaches. A computational example is presented which demonstrates the advantages of the new algorithm.

1. INTRODUCTION

The computational efficiency of wavelet transforms and the desirable time-frequency localization properties of their basis functions have motivated intensive research in the area of wavelet-based noise reduction. Many of the investigated approaches resemble principal-component analysis, in that the observed signal is projected onto a specially chosen set of basis functions, and expansion coefficients thought to represent noise are attenuated. One promising approach explored independently by Saito [1] and by Pesquet, *et al.* [2] uses the wavelet-packet bases of Coifman and Wickerhauser [3] and the Minimum Description Length (MDL) model selection criterion of Rissanen [4] to enhance signals in additive white noise.

In the MDL framework for noise reduction, a noisy observation $\boldsymbol{x} \in \boldsymbol{R}^N$ (consisting of a signal $\boldsymbol{s}$ in additive noise $\boldsymbol{n}$) is modeled as an output symbol from a discrete memoryless information source. The description length of $\boldsymbol{x}$, defined as the length (in bits) of a theoretical binary codeword used to describe $\boldsymbol{x}$, is given as:

$$L(\boldsymbol{x}, \lambda^{(k)}) = L(\lambda^{(k)}) + L(\boldsymbol{x}|\lambda^{(k)}) \quad (1)$$

where $L(\lambda^{(k)})$ and $L(\boldsymbol{x}|\lambda^{(k)})$ are the lengths of codewords respectively describing $\lambda^{(k)}$, a k-th order parametric model of $\boldsymbol{x}$, and the prediction error for the estimate $\hat{\boldsymbol{x}}(\lambda^{(k)})$ derived from the parametric model. $L(\lambda^{(k)})$ and $L(\boldsymbol{x}|\lambda^{(k)})$ are respectively determined by a universal prefix coding method proposed by Rissanen and by the Shannon coding method. Among admissible parametric models, the model which produces the minimum description length is selected as the model most representative of the signal.

In [1] and [2], the parameter vector $\lambda^{(k)}$ contained $N-k$ zeros and k transform coefficients. The k transform coefficients corresponded to the k of N orthonormal wavelet-packet basis vectors used to estimate $\boldsymbol{s}$. For this representation, Saito showed that

$$\lambda^{(k)} = \left\{\lambda : \max_{\lambda \in \Lambda(k)} \log_2 p_n(\boldsymbol{x} - \boldsymbol{\Phi}\lambda)\right\}, \quad (2)$$

$$L(\lambda^{(k)}) = \frac{3k}{2}\log_2 N + C_1, \quad (3)$$

and

$$L(\boldsymbol{x}|\lambda^{(k)}) = -\log_2 p_n(\hat{\boldsymbol{n}}(k)), \quad (4)$$

where $p_n(\boldsymbol{n})$ was the probability density of the noise, $\boldsymbol{\Phi}$ was an orthogonal matrix whose columns $\{\boldsymbol{\phi}_i\}_{i=1}^N$ were the wavelet-packet basis vectors used to describe $\boldsymbol{x}$, $\Lambda(k)$ was the subset of vectors in $\boldsymbol{R}^N$ with $N-k$ zero elements, $\hat{\boldsymbol{n}}(k) \triangleq \boldsymbol{x} - \boldsymbol{\Phi}\lambda^{(k)}$ was the implicit noise estimate corresponding to the kth-order signal model, and C_1 was a constant independent of basis or model order. For the case of white Gaussian noise with mean of zero and variance of σ_u^2,

$$L(\boldsymbol{x}|\lambda^{(k)}) = \hat{L}(\boldsymbol{x}|\lambda^{(k)}) + C_2 = \frac{\|\boldsymbol{x}\|^2 - \sum_{i=1}^{k}(\boldsymbol{\phi}_{z_i}{}^H\boldsymbol{x})^2}{2\sigma_u^2 \ln 2} + C_2 \quad (5)$$

where the index set $\{z_i\}_{i=1}^k$ denotes the largest k elements of $\boldsymbol{\Phi}^H\boldsymbol{x}$, and C_2 is a constant independent of basis or model order. The algorithms of [1, 2] found the value of k which minimized the sum of Equations 3 and 4, and returned a linear combination of the k wavelet-packet basis vectors with the largest expansion coefficients as a model of the original signal.

The MDL approach described above works well for the case of signals corrupted by additive white Gaussian noise. However, the restriction of the MDL approach to reduction of white noise limits its appropriateness for some practical applications.

* The work of N. Whitmal and J. Rutledge was supported in part by the National Science Foundation under grant BCS-9110247 and the Buehler Center on Aging. N. Whitmal was further supported by the Motorola Corporation and a Monsanto Foundation Faculty Internship.

† The work of J. Cohen was supported in part by a DePaul University Summer Research Grant.

2. THEORY

Autoregressive (AR) models provide a useful tool for adapting the MDL algorithm to reduction of correlated noise. A straightforward adaptation involves fitting an AR model to the noise component, building an FIR prediction-error filter from the AR model, and using the filter to whiten the noise component of $\boldsymbol{x}$. The filtered signal may then be denoised as described in Section 1.

In applications where denoised coefficients may be used as input parameters for other algorithms (e.g., speech enhancement and recognition), the straightforward approach becomes cumbersome. There, one is required either to inverse-transform, inverse-filter, and retransform the denoised signal, or to require the other algorithms to compensate for the pre-filtering operation. A new approach which selects the appropriate basis vectors in the original (unfiltered) domain is preferred.

The proposed approach (which is in part motivated by [5]) uses the wavelet-packet transform of a pth-order prediction error filter to reduce the correlated noise reduction problem to that of linear least-squares estimation. This approach *implicitly* transforms the original basis vectors into a second set of basis vectors which facilitates use of the MDL criterion in the original domain. We then use the MDL criterion to select the number k, the index set $\{z_i\}_{i=1}^{k}$, and coefficients $\lambda^{(k)}$ of the subset of basis vectors giving the best estimate of the original signal.

We assume that the original signal $\boldsymbol{s}$ may be represented by a subset of N periodized wavelet-packet basis vectors, and that the N elements $\{n_\ell\}_{\ell=0}^{N-1}$ of additive noise signal $\boldsymbol{n}$ are generated by an autoregressive random process of the form

$$n_\ell = -\sum_{m=1}^{p} \alpha_m n_{\ell-m} + u_\ell \tag{6}$$

where $p \ll N$, and $\{u_\ell\}$ is a Gaussian white noise process with mean of zero and variance of σ_u^2. For the model of Eq. 6, the probability density of $\boldsymbol{n}$ is [6]

$$p_n(\boldsymbol{n}) = \left((2\pi\sigma_u^2)^N \det \boldsymbol{C}_{\bar{n}}\right)^{-\frac{1}{2}} \exp\left[-\frac{Q(\boldsymbol{n})}{2\sigma_u^2}\right] \tag{7}$$

where

$$Q(\boldsymbol{n}) = \bar{\boldsymbol{n}}^H \boldsymbol{C}_{\bar{n}}^{-1} \bar{\boldsymbol{n}} + \sum_{\ell=p}^{N-1} \left(\sum_{m=0}^{p} \alpha_m n_{l-m} \right)^2$$

with $\alpha_0 = 1$, $\bar{\boldsymbol{n}} = [n_0\ n_1 \cdots n_{p-1}]^H$, and $\boldsymbol{C}_{\bar{n}} = E\{\bar{\boldsymbol{n}}\bar{\boldsymbol{n}}^H\}$. The codeword length for the prediction error of the kth order estimate of the signal is then given by

$$L(\boldsymbol{x}|\lambda^{(k)}) = \frac{N \ln(2\pi\sigma_u^2) + \ln \det \boldsymbol{C}_{\bar{n}} + Q(\boldsymbol{n})/\sigma_u^2}{2 \ln 2} \tag{8}$$

After substituting $\hat{\boldsymbol{n}}(k)$ for $\boldsymbol{n}$ and ignoring negligible [6] terms of Eq. 8, the *variable* portion of $L(\boldsymbol{x}|\lambda^{(k)})$ becomes

$$\hat{L}(\boldsymbol{x}|\lambda^{(k)}) \approx \frac{\|\boldsymbol{A}^H \boldsymbol{\Phi}_1(k)\lambda_1^{(k)} + \boldsymbol{A}^H \boldsymbol{\Phi}_2(k)\lambda_2^{(k)}\|^2}{2\sigma_u^2 \ln 2} \tag{9}$$

where

$$\boldsymbol{A}^H = \begin{bmatrix} \alpha_p & \alpha_{p-1} & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \alpha_p & \cdots & \alpha_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & \alpha_2 & \alpha_1 & 1 \end{bmatrix},$$

$$\boldsymbol{\Phi}_1(k) = \left[\ \boldsymbol{\phi}_{z_1} \quad \boldsymbol{\phi}_{z_2} \quad \cdots \quad \boldsymbol{\phi}_{z_k}\ \right],$$

$$\boldsymbol{\Phi}_2(k) = \left[\ \boldsymbol{\phi}_{z_k+1} \quad \boldsymbol{\phi}_{z_k+2} \quad \cdots \quad \boldsymbol{\phi}_{z_N}\ \right],$$

and $\lambda_i^{(k)} = \boldsymbol{\Phi}_i(k)^H \hat{\boldsymbol{n}}(k)$ for $i \in \{1,2\}$. Since the subspace spanned by $\{\boldsymbol{\phi}_{z_i}\}_{i=k+1}^{N}$ is assumed to only contain noise, $\lambda_2^{(k)} = \boldsymbol{\Phi}_2(k)^H \boldsymbol{x}$. The minimizing value of $\lambda_1^{(k)}$ may then be found by solving a set of normal equations, leading minimization of $\hat{L}(\boldsymbol{x}|\lambda^{(k)})$ to become equivalent to maximization of $\|\boldsymbol{A}^H \boldsymbol{\Phi}_1(k) \boldsymbol{\Phi}_1{}^H(k)\boldsymbol{x}\|^2$.

We now use the properties of the shift-invariant wavelet-packet transform to simplify the maximization problem. First, we note that the filtering matrix $\boldsymbol{A}^H$ consists of $N-p$ circular shifts of the prediction error filter, and equals

$$\boldsymbol{A}^H = \boldsymbol{S}\boldsymbol{F}\boldsymbol{Q}_A\boldsymbol{F}^H, \tag{10}$$

where $[\boldsymbol{F}]_{mn} = \frac{e^{j2\pi mn/N}}{\sqrt{N}}$, $\boldsymbol{S} = \left[\boldsymbol{I}_{(N-p)\times(N-p)} \vdots \boldsymbol{0}_{(N-p)\times p}\right]$, and $\boldsymbol{Q}_A$ is a diagonal matrix containing the values of the FFT of the first row of $\boldsymbol{A}^H$ divided by $\sqrt{N}$. We may then write

$$\boldsymbol{\Phi}^H \boldsymbol{A}\boldsymbol{A}^H \boldsymbol{\Phi} = \boldsymbol{\Phi}^H \boldsymbol{F}\boldsymbol{Q}_A^H \boldsymbol{Q}_A \boldsymbol{F}^H \boldsymbol{\Phi} - \boldsymbol{\Phi}^H \boldsymbol{F}\boldsymbol{\Theta}\boldsymbol{F}^H \boldsymbol{\Phi} \tag{11}$$

where $[\boldsymbol{\Theta}]_{mn} = q_m^* q_n \frac{\sin \frac{p\pi}{N}(m-n)}{N \sin \frac{\pi}{N}(m-n)} e^{j\pi(p+1)(m-n)/N}$ and $q_n \equiv [\boldsymbol{Q}_A]_{nn}$. Sinha and Tewfik [7, p. 3476] have shown that, for periodized Daubechies wavelets with large numbers of vanishing moments, $\boldsymbol{\Phi}^H \boldsymbol{F}\boldsymbol{Q}_A^H \boldsymbol{Q}_A \boldsymbol{F}^H \boldsymbol{\Phi}$ is nearly diagonal. And, since

$$\left|[\boldsymbol{\Theta}]_{mn}\right|^2 \le \frac{p^2}{N^4} \left(1 + \sum_{\ell=1}^{p} \alpha_\ell^2 \right)^2, \tag{12}$$

we may use the Cauchy-Schwarz inequality to show that each term in $\boldsymbol{\Phi}^H \boldsymbol{F}\boldsymbol{\Theta}\boldsymbol{F}^H \boldsymbol{\Phi}$ is at least a factor of $\frac{p}{N}$ less than those on the diagonal of $\boldsymbol{\Phi}^H \boldsymbol{F}\boldsymbol{Q}_A^H \boldsymbol{Q}_A \boldsymbol{F}^H \boldsymbol{\Phi}$. Hence, the columns of $\boldsymbol{A}^H \boldsymbol{\Phi}_1(k)$ are nearly orthogonal,

$$\|\boldsymbol{A}^H \boldsymbol{\Phi}_1(k) \boldsymbol{\Phi}_1{}^H(k)\boldsymbol{x}\|^2 \approx \sum_{i=1}^{k} (\boldsymbol{\phi}_{z_i}{}^H \boldsymbol{x})^2 \|\boldsymbol{A}^H \boldsymbol{\phi}_{z_i}\|^2, \tag{13}$$

and

$$\hat{L}(\boldsymbol{x}|\lambda^{(k)}) \approx \frac{\|\boldsymbol{A}^H \boldsymbol{x}\|^2 - \sum_{i=1}^{k} (\boldsymbol{\phi}_{z_i}{}^H \boldsymbol{x})^2 \|\boldsymbol{A}^H \boldsymbol{\phi}_{z_i}\|^2}{2\sigma_u^2 \ln 2}, \tag{14}$$

where the index set $\{z_i\}_{i=1}^{k}$ corresponds to the set of *filtered* basis vectors with the largest coefficients. Equation 14 is similar to Eq. 5 of the case of white noise, with the transformed basis system $\boldsymbol{A}^H \boldsymbol{\Phi}$ used in place of the original basis system.

For a second simplification, we note that the columns of the matrix $A^H\Phi$ are projections of $N-p$ shifted versions of a time-reversed prediction error filter onto the "signal" subspace. Calculation of these transform coefficients may be done quickly by means of the shift-invariant wavelet-packet transform [2], which, by calculating the transforms of all circulant shifts of a given signal, may be used to *convolve* the basis functions with any chosen filter. Hence, a fast algorithm similar to those of [1, 2] may be used to select those basis vectors which best represent the signal.

A further increase in efficiency may be gained by following an approach of Krim and Pesquet [8]. By taking the finite difference (with respect to k) of the sum of Equations 3 and 14, and looking for the value of k for which that difference changes from negative to positive, the MDL criterion may be shown to select the largest value of k for which

$$|\phi_{z_k}{}^H x| > \frac{\sigma_u}{\|A^H\phi_{z_k}\|}\sqrt{2.0794\log_2 N}.$$

Evaluation of $L(x|\lambda^{(k)})$ for various values of k may then be replaced by a simple thresholding operation, similar to that of [2], [8], and [9]. In contrast to these approaches, the new method provides basis-dependent thresholding which is more appropriately tailored to the characteristics of the noise.

3. BASIS SELECTION

The proposed algorithm also uses the local discrimination bases (LDBs) developed by Saito and Coifman [10] as a feature extraction tool for pattern recognition. In the LDB formulation, the statistical properties of each of C signal classes are described by C time-frequency-energy (TFE) maps constructed from sets of training data. For class $\ell \in \{1,2,\ldots,C\}$ with N_ℓ training signals $\{x_\ell(i)\}_{i=1}^{N_\ell}$, the TFE map $\Gamma_\ell(j,k,m)$ was defined as

$$\Gamma_\ell(j,k,m) = \left[\sum_{i=1}^{N_\ell}|\phi_{j,k,m}{}^H x_\ell(i)|^2 \Big/ \sum_{i=1}^{N_\ell}\|x_\ell(i)\|^2\right] \quad (15)$$

where $\phi_{j,k,m}$ is the wavelet-packet basis vector of scale j, frequency band k, and position m in the binary tree. For discrimination between the classes, Saito used a "best-basis" algorithm [3] to select the basis partition which maximized the Kullback-Liebler distance

$$D(\{p\},\{q\}) = \sum_i p_i \log_2 \frac{p_i}{q_i} \quad (16)$$

between the "probability" (i.e., TFE map value) distributions $\{p_i\}$ and $\{q_i\}$ of each pairwise combination of each of the C classes. Saito applied his algorithm to the denoising problem by selecting bases which maximized discrimination between the two classes of "signal+noise" ($\{p_i\}$) and "noise" ($\{q_i\}$).

In our extension of the LDB approach, we note that the expression of Eq. 15 is an approximation to the ratio of expected energies

$$\Gamma_\ell(j,k,m) = E\{|\phi_{j,k,m}{}^H x_\ell|^2\}/E\{\|x_\ell\|^2\}. \quad (17)$$

We therefore propose a new approach for calculating the TFE map of the "noise" class which does not require prior training. The new approach requires only the calculation of the shift-invariant transform (into the smallest subbands) of the impulse response of the AR filter. At each level of the transform, the energy of the transform coefficients are summed to calculate the numerator term of Eq. 17.

4. SUMMARY OF ALGORITHM

We now summarize the sequence of operations in our algorithm:

1. Compute the shift-invariant wavelet-packet transforms of the signal and (as needed) the filter response.
2. Construct TFE maps for the "signal+noise" class and (as needed) the "noise" class.
3. Use the TFE maps of Step 2 to find the best LDB partition for denoising.
4. Use one of the two new denoising approaches developed in Section 2 to implement shift-invariant denoising.
5. If desired, use the inverse shift-invariant wavelet-packet transform to resynthesize the signal.

5. EXPERIMENTAL RESULTS

One objective of our current research is the development of a new noise reduction method which can enhance noisy speech signals without impairing speech intelligibility. Towards this end, we evaluated the new approach (with thresholding) for use in a practical speech enhancement algorithm which implements denoising on a frame-by-frame basis. (Motivation for this approach is discussed in [11].) As input, we used a 130 ms recording of "/bi/" (preceded by 130 ms of silence) which was sampled at 16 kHz and added to 4096 samples of AR Gaussian noise of the form $n_\ell = 0.8n_{\ell-1} + u_\ell$. The variance of $\{u_\ell\}$ was selected to give a speech-to-noise ratio of 5 dB. Both clean and noisy versions of the signal are shown in Figure 1.

Estimates of the noise parameters were taken from the first 512-sample analysis frame (located in the silence region) and used in the denoising of subsequent frames. AR models with $p=1$ and $p=0$ were fit to the first frame's data, with the latter corresponding to the standard white noise case explored in [1, 2, 8, 9]. TFE maps for the "speech+noise" class were calculated over a range of five circulant shifts and averaged for use in the algorithm.

Output waveforms for both cases (using the Daubechies-20 wavelet) are shown in Figure 2. Noise suppression and SNR measures for (respectively) the silence and speech regions are also shown below in Table 1. Both sets of data indicate that the new method enjoys a significant advantage in performance, particularly in the silence regions. Recent results [12] indicate that the advantage in silence may be useful in enhancing the perceived quality of the noisy speech. Informal listening also indicates that the new algorithm's output sounds more natural than that of the other algorithm (which tends to sound muffled).

We note here that the advantage of our algorithm is less pronounced in the speech region. This is in part caused by the prediction-error filter, which tends to partially decorrelate the speech and make some speech components subject to attenuation by thresholding. It should also be noted that for longer analysis frames, the combined use of the LDB approach and shift-invariant transforms can result in transform structures with impractical computation requirements. Our present work is focused on development of an intelligibility-based LDB algorithm which addresses both of these concerns.

6. CONCLUSION

We have presented a new wavelet-based method for reducing additive autoregressive noise. The new method uses the shift-invariant wavelet-packet transform to implement basis-dependent thresholding. Preliminary results suggest that the new method could be particularly useful in enhancing noisy speech signals.

7. REFERENCES

[1] N. Saito, "Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion," in: *Wavelets in Geophysics*, E. Foufoula-Georgiou and P. Kumar, eds.: Academic Press, 1994.

[2] J.-C. Pesquet, H. Krim, H. Carfantan, and J.G. Proakis, "Estimation of noisy signals using time-invariant wavelet packets," 27th Asilomar Conf. Sig. Syst. Comp., pp. 635-640, 1993.

[3] R.R. Coifman and M.V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inf. Theory*, vol. 38, pp. 713-718, 1992.

[4] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*: Singapore, World Scientific Publishing, 1989.

[5] S. Kay and V. Nagesha, "Maximum likelihood estimation of signals in autoregressive noise," *IEEE Trans. Sig. Proc.*, vol. 42, pp. 88-101, 1994.

[6] G.E.P. Box and G.M. Jenkins, *Time Series Analysis – Forecasting and Control*: San Francisco, Holden Day, 1970.

[7] D. Sinha and A.H. Tewfik, "Low bit rate transparent audio compression using adapted wavelets," *IEEE Trans. Sig. Proc.*, vol. 41, p. 3463-3479, 1993.

[8] H. Krim and J.-C. Pesquet, "On the Statistics of Best Bases Criteria," in: *Wavelets and Statistics*, A. Antoniadis, ed.: Springer-Verlag, 1995.

[9] D.L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, pp. 613-627, 1995.

[10] N. Saito and R.R. Coifman, "Local discriminant bases and their applications," *Jour. Math. Imag. Vision*, vol. 5, pp. 337-358, 1995.

[11] J. Berger, R.R. Coifman, and M. Goldberg, "Removing noise from music using local trigonometric bases and wavelet packets," *Jour. Audio Eng. Soc.*, vol. 42, pp. 808-818, 1994.

[12] R. Tyler and T.J. Johnson, "Attenuating Noise During Pauses in Speech," *NIH Interdisciplinary Forum on Hearing Aid Research and Development*, September, 1995.

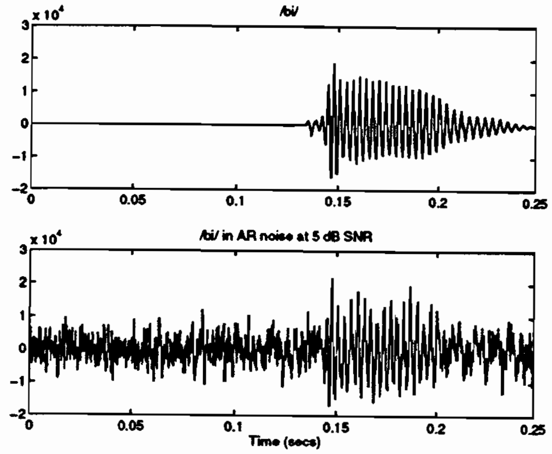

Figure 1: Clean and noisy versions of /bi/

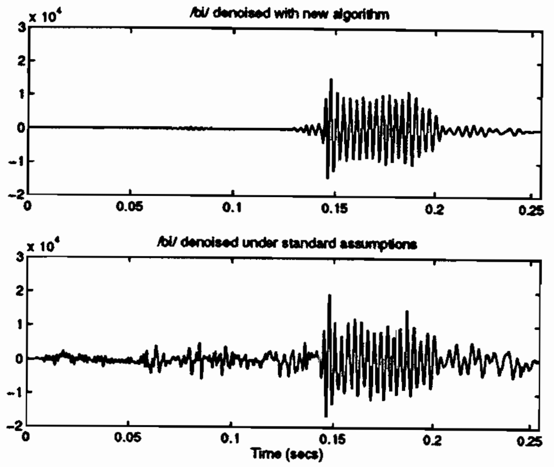

Figure 2: Denoised versions of /bi/

Table 1: Objective measures of algorithm performance

	Reduction of noise (silence region)	Enhancement capability (speech region)
New method	32.98 dB	SNR = 9.32 dB
Old method	7.39 dB	SNR = 7.78 dB