A Single-Channel ICA-R Method for Speech Signal Denoising combining EMD and Wavelet

Yangyang Qi

Institute of Communication Engineering, PLA Univ. of Sci.&Tech, Nanjing, China
Email: jimmy_qi@yeah.net

Fuqiang Yao and Miao Yu

Nanjing Telecommunication Technology Research Institute, Nanjing, China

Abstract—**According to the problem of speech signal denoising, we propose a novel method in this paper, which combines empirical mode decomposition (EMD), wavelet threshold denoising and independent component analysis with reference (ICA-R). Because there is only one mixed recording, it is a single-channel independent component analysis (SCICA) problem in fact, which is hard to solve by traditional ICA methods. EMD is exploited to expand the single-channel received signal into several intrinsic mode functions (IMFs) in advance, therefore traditional ICA of multi-dimension becomes applicable. First, the received signal is segmented to reduce the processing delay. Secondly, wavelet thresholding is applied to the noise-dominated IMFs. Finally, fast ICA-R is introduced to extract the object speech component from the processed IMFs, whose reference signal is constructed by assembling the high-order IMFs. The simulations are carried out under different noise levels and the performance of the proposed method is compared with EMD, wavelet thresholding, EMD-wavelet and EMD-ICA approaches. Simulation results indicate that the proposed method exhibit superior denoising performance especially when signal-to-noise ratio is low, with a half shorter running time.**

Index Terms—**speech signal denoising, EMD, wavelet, independent component analysis; SCICA; fast ICA-R**

I. INTRODUCTION

Speech signal denoising is a classic problem in signal processing. Assuming we observe a noisy such that

$$x(t) = s(t) + \sigma n(t) . \qquad (1)$$

where $s(t)$ is the original speech signal, $n(t)$ is Gaussian white noise whose statistical distribution obeys $\mathcal{N}(0,1)$, σ is the variance. Our goal is to obtain a denoised estimation $\hat{s}(t)$, using only the statistical properties of $x(t)$.

To settle such problem, linear filters such as the Wiener filtering are frequently used because they are easy to implement. However, these linear methods are not effective when σ is unknown [1]. According to this, other approaches have been proposed, especially those based on empirical mode decomposition (EMD), wavelet thresholding and independent component analysis (ICA).

EMD was proposed by N. E. Huang in 1998 [2]. EMD decomposes the signal into several intrinsic mode functions (IMFs) through an iterative process called sifting. A generalized task for EMD is signal denoising, which is accomplished by reconstructing the signal with the IMFs containing useful information, abandoning the noise-dominated ones. [1]. Hence, how to choose the proper number of noise-dominated IMFs is an important factor in EMD approach, which affects denoising performance greatly.

With respect to wavelet approach, the signal is transformed into wavelet domain at first. The energy of the speech signal often focuses in a few wavelet components with high amplitudes, while the energy of noise spreads over all coefficients with low amplitudes [3]. The wavelet coefficients are compared to a given threshold value and then they are modified depending on the thresholding rule. Finally, inverse wavelet transform is performed to obtain the denoised signal. A main drawback of the wavelet approach is that the basic functions are fixed, which cannot match the varying nature of signals [3]

In addition, simply EMD approach or wavelet thresholding cannot achieve satisfactory performance when the signal-to-noise-ratio (SNR) is low [4]. Therefore a combination method of EMD and wavelet is introduced, which can be termed as EMD-wavelet [5].

Independent component analysis (ICA) is a famous approach for blind source separation (BSS) problems. ICA is dedicated to recover a set of unknown mutually independent source signals from their observed mixtures without prior-knowledge of the mixing coefficients [6]. If the speech signal can be extracted from the noisy mixture by ICA, then great improvement in SNR will be achieved [7].

Traditional ICA is only capable of tackling the problem when the number of channels is larger than or equal to the number of sources. When there is only one sensor in the receiver end, the problem comes into single-channel independent component analysis (SCICA) [8]. Because SCICA is the extreme case of underdetermined

problem, traditional ICA of multi-dimension becomes incompetent.

An approach combining EMD and ICA is proposed in [9], which attempts to deal with the SCICA problems. This approach is described as EMD-ICA. The main shortcoming of EMD-ICA denoising approach is that the constructed virtual noise channel can't necessarily match with the real white one [10].

Besides, traditional ICA has to recover all the source signals, which causes the trouble of selection. However, ICA with reference (ICA-R) is able to extract the desired component from mixture when the pre-defined reference signal is available [11]. But the adoption of ICA-R results in heavy computation load. Moreover, the reference signal is hard to construct when the background noise is very strong.

We proposed a basic idea of speech signal denosing which combines EMD-wavelet and ICA-R in [12]. In this paper, we further this work by replacing original ICA-R algorithm by fast ICA-R. In addition, according to the nonlinear relationship between the length of the data and the running time of EMD algorithm, we segment the noisy speech in advance to reduce the processing delay. What's more, EMD-ICA denoising method is brought into study and we make a comprehensive comparison of all the approaches aforementioned. Simulations were performed to verify the validity of the proposed method. The detailed analyses of the results demonstrate the novel method proposed in this paper exhibits superior denoising performance with a half shorter running time.

II. Basic Theory

A. Emprical Mode Decompositon

EMD decomposes a given signal into a series of IMFs one after another, starting from high frequency to low frequency [13]. In contrast to conventional decomposition methods, the basic functions of EMD are derived from the signal itself rather than defined in advance [3].

An IMF is defined as a function with equal number of extrema and zero crossings (or at most differed by one) with its envelopes, as defined by all the local maxima and minima, being symmetric with respect to zero [13]. The total sum of the IMFs matches the signal very well and therefore ensures completeness.

$$x(t) = \sum_{i=1}^{n} imf_i(t) . \qquad (2)$$

For convenience, we refer to $imf_i(t)$ as the i th-order IMF. By this convention, lower order IMFs capture high frequency modes while higher order IMFs typically represent low frequency modes.

Considering the IMFs acquired from the decomposition of noise speech signal, lower order IMFs comprises the large amount of noise compared to the rest ones, so a denoised signal can be acquired by reconstructing the signal using only the high-order IMFs. However, EMD denoising principle may cause great distortion because the noise-dominated IMFs are likely to

contain some signal portions as well, thus the signals' structures or features cannot be well preserved [4].

Although EMD often results in useful decomposition outcomes, it is not mathematical established [5]. Some inferences in the paper are based on large amount of simulations other than sound mathematically theory.

B. Wavelet Thresholing

Discrete Wavelet Transform (DWT) projects the signal onto a number of predefined orthonormal basis functions, which are termed mother wavelet [4]. Then signal is expressed by a linear sum of the mother wavelet and wavelet coefficients. Therefore, DWT may not match the real signal if the pre-selected mother wavelet is not appropriate [3].

Wavelet thresholding can be briefly described as follows. DWT transforms signal $\mathbf{x} = [x_1, x_2, ..., x_N]$ into wavelet domain by a orthogonal $N \times N$ transform matrix $\mathbf{U}$,

$$\mathbf{c} = \mathbf{U}\mathbf{x} . \qquad (3)$$

where $\mathbf{c} = [c_1, c_2, ... c_N]$ represents the wavelet coefficients.

There are two major thresholding rules, hard and soft thresholding, which can be referred [5]

$$\zeta_T(A) = \begin{cases} A, & |A| > T \\ 0, & |A| \le T \end{cases} . \qquad (4)$$

$$\zeta_T(A) = \begin{cases} \text{sgn}(A)(|A| - T), & |A| > T \\ 0, & |A| \le T \end{cases} . \qquad (5)$$

where A is the amplitude of wavelet coefficient, $\zeta_T(A)$ is the amplitude after thresholding and T is the thresholding value. Soft thresholding cause no discontinuities in the denoised signal, so it is preferred in practice for speech denoising.

Using any one of the thresholoding rule above, the denoised signal can be estimated by inverse discrete wavelet transform

$$\hat{\mathbf{x}} = \mathbf{U^T}\tilde{\mathbf{c}} . \qquad (6)$$

where $\tilde{\mathbf{c}} = [\zeta_T(c_1), \zeta_T(c_2), ... \zeta_T(c_N)]$, $\mathbf{U^T}$ represents the transposition of matrix $\mathbf{U}$ and $\hat{\mathbf{x}} = [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_N]$ is an almost noise-free version of $\mathbf{x}$.

The performance of denoising in wavelet domain depends on the threshold value particularly. The most widely used thresholding method is VisuShrink [14], which is given by

$$T = \sigma\sqrt{2\log N} . \qquad (7)$$

where σ can be estimated based on the component median [14]

$$\hat{\sigma} = \frac{\text{median}(|c_i| : i = 1, ... N)}{0.6745} \qquad (8)$$

However, research indicated that simply wavelet thresholding is not ideal when SNR is low [5].

C. EMD-wavelet

EMD-wavelet method is proposed to obtain a denoised signal with minimal distortion.

Firstly, the noisy signal is decomposed into several IMFs by EMD adaptively. Then wavelet thresholding is applied to all IMFs to exclude components that are expected to be significantly corrupted by noise. Finally, the signal is constructed by the processed IMFs.

The soft thresholding methods can be translated to [5]

$$\widetilde{imf}_i(t) = \begin{cases} \text{sgn}(imf_i(t))(|imf_i(t)| - T), & |imf_i(t)| > T \\ 0, & |imf_i(t)| \le T \end{cases} \quad (9)$$

where $\widetilde{imf}_i(t)$ indicates the ith processed IMF.

The denoised signal can be written by

$$\hat{x}(t) = \sum_{i=1}^{n} \widetilde{imf}_i(t) \quad (10)$$

In EMD-wavelet method, EMD is used to analysis the noisy signal other than wavelet decomposition, so it is dispensable to determine the basis functions by exploiting prior-knowledge. At the same time, IMFs are "shrinked" other than being discarded completely, which results in a relatively less distortion signal [15].

D. Independent Component Ananlysis

Independent component analysis (ICA) is a statistical method for transforming an observed multi-dimensional random vector into components that are statistically as independent from each other as possible [16]. In 1994, P. Comon theorized ICA as the process of dynamic optimization (learning rule) under certain object (contrast) function [17]. Negentropy, introduced by A. Hyvärinen in [6], is largely used as the object function. Experimentally speaking, the ICA algorithm based on negentropy always works better than that using other object functions [18].

Considering the scenario that the source signals are linearly mixed instantaneously, the received signals can be represented as

$$\mathbf{M} = \mathbf{AS} \quad (11)$$

where $\mathbf{M} = [m_1, m_2, \cdots, m_p]$ is the aggregation of the received signal, $\mathbf{S} = [s_1, s_2, \cdots, s_q]$ is the aggregation of the source signal, $\mathbf{A}$ is the mixing matrix. If the inverse of $\mathbf{A}$, has been estimated accurately by ICA, supposing it to be $\mathbf{W} = \mathbf{A}^{-1}$, then the source signals could be recovered by

$$\hat{\mathbf{S}} = \mathbf{W} \cdot \mathbf{M} = \mathbf{W} \cdot (\mathbf{AS}) = \mathbf{S} \quad (12)$$

where $\hat{\mathbf{S}} = [\hat{s}_1, \hat{s}_2, \cdots, \hat{s}_q]$ is the estimation of $\mathbf{S}$.

For common ICA method, $p \ge q$ is needed. Single-channel ICA is such an extreme case that $q \ge 2$ and $p = 1$, which is impossible for common ICA in realization [8].

However, EMD brings a promising approach by expanding the single channel received signal into multi-channel IMFs, transforming the single channel underdetermined problem into the multi-channel positive definite problem. The approach is described as EMD-ICA [9].

With respect to signal denosing, EMD-ICA first constructs a virtual noise channel using low-order IMFs [10]. Assuming the noisy speech is $\mathbf{X}$, the virtual noise channel is $\mathbf{Y}$, the real noise is $\mathbf{n}$, speech signal is $\mathbf{S}$. Supposing the virtual noise channel draw very near to the real noise, namely $\mathbf{n} \approx \mathbf{n_1}$

$$\mathbf{X} = a_{11}\mathbf{n} + a_{12}\mathbf{S}$$
$$\mathbf{Y} = a_{21}\mathbf{n_1} \approx a_{21}\mathbf{n} \quad (13)$$

where a_{ij} i, j =1,2 are mixing coefficients. Then, (13) can be written in a matrix form

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{n} \\ \mathbf{S} \end{pmatrix} = \mathbf{A} \begin{pmatrix} \mathbf{n} \\ \mathbf{S} \end{pmatrix} \quad (14)$$

where $\mathbf{A}$ is the mixing matrix.

If $\mathbf{B}$ is the estimated inversion of $\mathbf{A}$, supposing $\mathbf{B} \approx \mathbf{A}^{-1}$, then the source signals could be recovered by

$$\begin{pmatrix} \mathbf{n} \\ \mathbf{S} \end{pmatrix} = \mathbf{B} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \quad (15)$$

But in fact the noise contained in each IMF is colored, having different energy in each mode, so the virtual channel can't necessarily match with the real white one [10].

It is also noticeable that the common ICA is only capable of recover all the independent components, because the whole unmixing matrix $\mathbf{W}$ is present in the learning rule [11]. This is inefficient in some application scenario, especially where only few source signals are desired. Although some ICA algorithms are able to extract ICs one at a time with a deflation process, the arbitrary order of extraction remains as the major drawback [19].

The ICA with reference (ICA-R) is a good candidate to solve this problem when the pre-defined reference signals are available. The framework of original ICA-R can be formulated as follows [11]

$$\text{maximize} \quad J(y) \approx \rho \left[E\{G(y)\} - E\{G(v)\} \right]^2$$
$$\text{subject to} \quad g(\mathbf{w}) \le 0, h(\mathbf{w}) = E\{y^2\} - 1 = 0 \quad (16)$$

where $J(y)$ is a reliable approximation of negentropy, being defined as the estimator for independent in this framework. The inequality constraint $g(\mathbf{w})$ is exploited to incorporate prior information of the desired signal into the ICA learning rule. The equality constraint $h(\mathbf{w})$ is used to bound $J(y)$ and the weight vector $\mathbf{w}$.

And the augmented Lagrangian function $L(\mathbf{w}, \mu, \lambda)$ for (16) can be formulated as follows:

$$L(\mathbf{w},\mu,\lambda)=J(y)-\frac{1}{2\gamma}[\max^2\{\mu+\gamma g(\mathbf{w}),0\}-\mu^2]-\lambda h(\mathbf{w})-\frac{1}{2}\gamma\|h(\mathbf{w})\|^2 \tag{17}$$

where μ and λ are Lagrange multipliers for the constrains $g(\mathbf{w})$ and $h(\mathbf{w})$ respectively. γ is a scalar penalty parameter. $\|\cdot\|$ denotes the Euclidean norm. A Newton-like learning algorithm can be derived by finding the maximum of $L(\mathbf{w}, \mu, \lambda)$ as follows:

$$\mathbf{w}_{k+1}=\mathbf{w}_k-\eta\mathbf{R}_{\mathbf{xx}}^{-1}L'_{w_k}/\delta(\mathbf{w}_k) \tag{18}$$

where k is an iteration index, η is the learning rate, $\mathbf{R}_{\mathbf{xx}}$ is the covariance matrix of the observed signal, L'_{w_k} is the first derivative of $L(\mathbf{w}, \mu, \lambda)$ with respect to $\mathbf{w}$. The definition of $\delta(\mathbf{w}_k)$ is detailed in [11].

As can be seen in (18), the computation of the inversion of $\mathbf{R}_{\mathbf{xx}}$ is involved in the learning algorithm, which is time consuming therefore limits the application of ICA-R in practice. In order to cut down the computation complexity, fast ICA-R is adopted in our method.

The improvement is based on the following two considerations [19].

- By normalizing the weight vector $\mathbf{w}$, the equality constraint $h(\mathbf{w})$ and λ can be omitted. The corresponding augmented Lagrangian function for the fast ICA-R.

$$L(\mathbf{w},\mu)=J(y)-\frac{1}{2\gamma}\Big[\max^2\{\mu+\gamma g(\mathbf{w}),0\}-\mu^2\Big] \tag{19}$$

- By pre-whitening the observed signals, the covariance matrix of whiten signal equals unity. Then the computing of $\mathbf{R}_{\mathbf{xx}}^{-1}$ in (18) is avoided.

$$\mathbf{w}_{k+1}=\mathbf{w}_k-\eta L'_{w_k}/\delta(\mathbf{w}_k) \tag{20}$$

Then, weight vector is normalized by.

$$\mathbf{w}_{k+1}=\mathbf{w}_{k+1}/\|\mathbf{w}_{k+1}\| \tag{21}$$

As a result, the computation load for the fast ICA-R is considerably reduced.

Another limitation of ICA-R is how to construct the reference signal. Because the speech has pitch and its harmonic frequencies, periodic rectangular pulses with pitch frequency of the speech signal can be used as the reference signal [20]. But few methods have been proposed to determine the pitch frequency when the speech is buried in noise. According to this problem, a novel way to construct the reference signal for ICA-R is proposed in this paper. We get the reference signal by assembling high-order IMFs, which works well even if the speech signal is buried in noise.

III. ALGORITHM

The method we proposed incorporates the advantages of EMD, wavelet and ICA, so it is referred to as EMD-wavelet-ICA.

The algorithm can be stated as below:

- Step1. Decompose the contaminated signal by EMD into a set of IMFs.
- Step2. Apply wavelet thresholding to the frontal two IMFs to get two denoised IMFs, $\widetilde{imf}_1(t)$ and $\widetilde{imf}_2(t)$.
- Step3. Summing over the remaining IMFs to get a reference signal $r(t)$
- Step4. Extract the object signal $z(t)$ from $\widetilde{imf}_1(t)$, $\widetilde{imf}_2(t)$ and $imf_3(t)$ by fast ICA-R.
- Step5. Summing over $z(t)$ and $r(t)$ to get the denoised signal.

The block diagram of the algorithm is illustrated in Fig. 1. Time variable t is omitted in the figure for simplicity.

Note that in Step2, it is totally unnecessary to apply thresholding to all IMFs with respect to white Gaussian noise removal. Equation (22) displays the difference between our method and the conventional EMD-wavelet.

$$\hat{x}(t)=\sum_{i=1}^{c_1}\widetilde{imf}_i(t)+\sum_{c_2}^{n}imf_i(t) \tag{22}$$

where c_1 denotes the number of high-order IMFs to be processed by wavelet thresholding, and $c_2=c_1+1$. The introduction of parameters c_1 and c_2 gives us flexibility to process the noisy low-order IMFs and preserve the high-order ones that contains primarily speech signal.

Huang proves in [21] that at least in the noise-only case, the distribution of the IMF samples still obeys Gaussian distribution, while IMFs contains mainly speech components do not follow the certain characteristics. Taking advantage of this discrepancy, we set the auto-correlation of Gaussian white noise as the criterion and calculate the autocorrelation of all the IMFs.

$$R_{ii}(\gamma)=E\Big[imf_i(t)imf_i(t+\gamma)\Big] \tag{23}$$

where γ denotes shift. Then the results are compared with the criterion. Simulations have proved that the auto-correlation of frontal two IMFs always get a relatively large maximum when $\gamma=0$ but rather small values at the rest, so imf_1 and imf_2 have the maximum possibility to be the noise-dominated IMFs. Consequently, we set $c_1=2$, $c_2=3$ in the propose method.
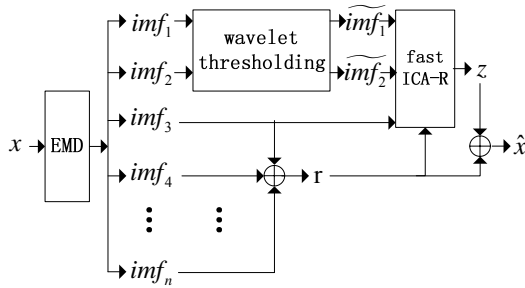
Figure 1.   The framework of the EMD-wavelet-ICA.

No noticeable degradation is observed in simulations by using this slightly modified version of EMD-wavelet compared to the original version.

It goes beyond doubt that the way we combine EMD and wavelet is not optimum. On the one hand, wavelet can be more appropriately adapted by exploiting the special characteristics of EMD, such as iterative EMD interval-thresholding, which effects better than the direct combination version [5]. On the other hands, although universal threshold ensures nearly no Gaussian noise will be left in the denoised signal, the elimination of possibly informative parts of signal is inevitable [14]. Some other threshold values are capable of preserving more useful information, such as the IMF-dependent threshold, characterized by adopting different thresholds T per mode [5]. But what we concentrate on in this paper is the exploration combining EMD-wavelet with ICA-R in single-channel speech signal denoising. Therefore, the direct combination version of EMD-wavelet and universal thresholding are adopted for simplicity.

In Step3, it is worthy to point out that EMD-wavelet is not exploited as a merely pre-processing tools, it becomes an integral part of the proposed method. Because the reference signal $r(t)$ we used in the following separation process is constructed by the remaining IMFs.

$$r(t) = \sum_{i=3}^{n} imf_i(t) \qquad (24)$$

Although lacking high frequency of the speech signal, it indeed exhibit satisfactory performance by demonstration of the simulation results. Then the denoised result can be written by

$$\hat{\mathrm{x}}(t) = \sum_{i=1}^{3} \widetilde{imf_i}(t) + r(t) \qquad (25)$$

where $\widetilde{imf_i}(t)$ indicates the speech component extracted from the frontal three IMFs.

In Step4, the employment of fast ICA-R not only makes the whole algorithm automatic and but also reduces the computation consumption of original ICA-R at the same time. Simulations on speech signal were also performed in [19], the results demonstrated the efficiency and accuracy of fast ICA-R.

In addition, there are various additional ways to combine EMD-wavelet with fast ICA-R. For example, fast ICA-R can be applied to all the IMF instead of merely focusing on the frontal three ones. But previous simulation result had proved that the approach focusing all the IMFs does not make distinct enhance performance [12].

IV. SIMUALATION RESULTS

To verify the validity of EMD-wavelet-ICA, simulations on the auto-correlation criterion, the output SNR, the correlation coefficient and the running time are performed. Then the results are analyzed and compared to the methods aforementioned.

A. *The Auto-correlation Criterion*

The auto-correlation of the Gaussian white noise and the frontal three IMFs from the decomposition of noisy speech when SNR=-3 dB are shown in Fig. 2.

In Fig. 2 (b) and (c), the auto-correlation properties of the frontal two IMFs indicate that they are noise-dominated. In Fig. 2 (d), the short-time auto-correlation properties of speech signal is dominated, which indicates imf_3 mainly contains speech signal.
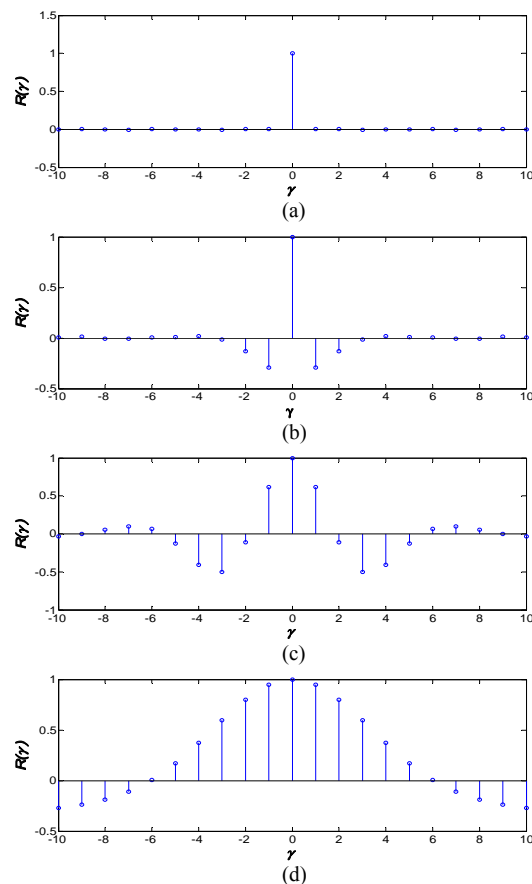

Figure 2.   (a)The auto-correlation of Gaussian white noise (b)The auto-correlation of imf_1 (c)The auto-correlation of imf_2 (d)The auto-correlation of imf_3

B. Waveform

The waveforms of the original speech signal, the noisy received signal corresponding to SNR=-3 dB and the denoised signals by different approaches are demonstrated in Fig. 4.

Fig. 4 (d) and (e) demonstrate that using different mother wavelet will lead to different denoising performance. It's obvious that db6 mother wavelet outperforms sym8 as for this chosen speech signal. In EMD-wavelet-ICA approach, the selection of mother wavelet is also unavoidable in Step2. However, EMD-wavelet-ICA utilizes EMD other than wavelet decomposition to analysis the signal in Step1, hence the influence of selecting appropriate mother wavelet is brought to minimum. For simplicity, we choose db6 mother wavelet in the following simulations.

Fig. 4 (c),(d),(e),(f) and (g) show that the denoised result by simply EMD or wavelet still contain considerable residual noise, while EMD-ICA and EMD-wavelet perform much better, but not extraordinary satisfactory.

As can be seen from Fig. 4, our method exhibits a superior performance in denoising.

C. Output SNR

To evaluate the performance of the denoising method, average values of the output SNR is calculated.

$$\text{SNR}_{output} = 10\log \frac{E[|s(t)|^2]}{E[|\hat{s}(t) - s(t)|^2]} \qquad (25)$$

where $s(t)$ and $\hat{s}(t)$ are the original and the estimated signals respectively. 100 independent noise sequences are generated for each input SNR value. The relationship between the average output SNR and the input SNR is shown in Fig. 3.

In Fig. 3, an improvement up to 7.2 dB is obtained by EMD-wavelet-ICA when input SNR=-4 dB. The output SNR curve of EMD-wavelet-ICA outperforms the other approaches, especially when the input SNR is low. In fact, it achieves nearly 0.7 dB gain against EMD-wavelet and even higher gain compared to other methods.
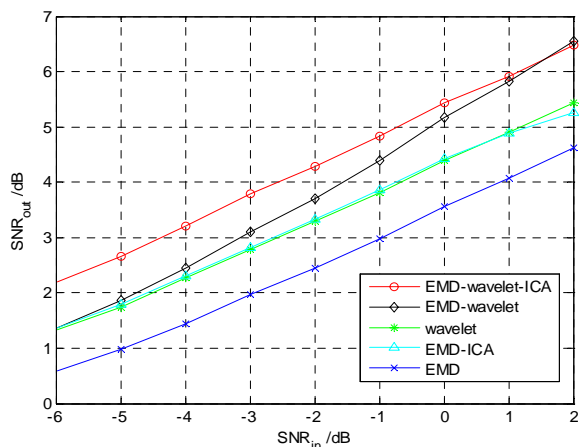

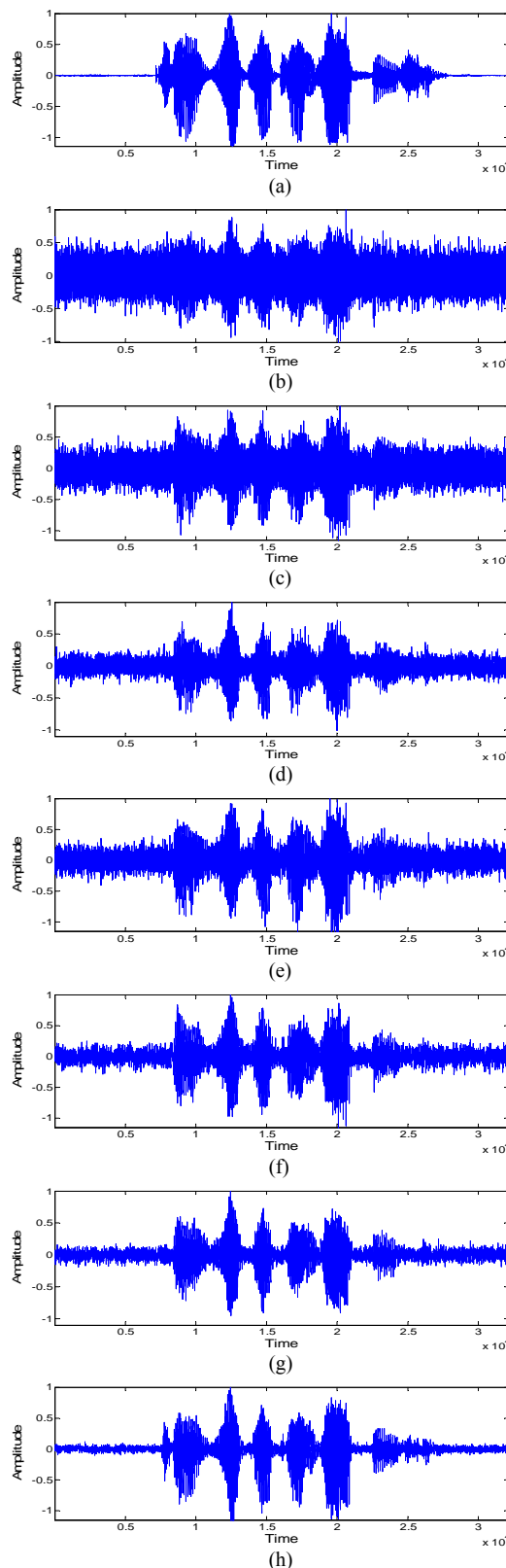Figure 3.   The relationship of output SNR and input SNR

Figure 4.   (a)The waveform of the clean speech signal (b)The waveform of the noisy signal corresponding to SNR=-3 dB (c)The waveform of the denoised result by EMD (d)The waveform of the denoised result by the wavelet (db6 mother wavelet) e)The waveform of the denoised result by the wavelet (sym8 mother wavelet) (f)The waveform of the denoised result by the EMD-ICA (g)The waveform of the denoised result by the EMD-wavelet (h)The waveform of the denoised result by the EMD-wavelet-ICA
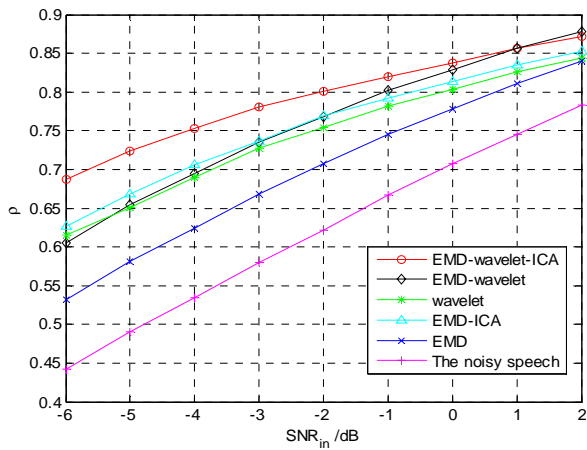
Figure 5. The relationship between correlation coefficient and input SNR

Figure 6. Performance comparison between fast ICA-R, ICA-R and FastICA algorithms

TABLE I.
THE RUNNING TIME OF THE APPROACHES AFOREMENTIONED

τ_p /s	SNR_{in} /dB				
	-6	-4	-2	0	2
EMD	4.2610	4.2613.	4.2410.	4.2313.	4.2184
Wavelet	0. 1352	0. 1333	0.1402	0.1394	0.1386
EMD-wavelet	4.3962	4.3968	4.3825	4.3683	4.3533
EMD-ICA	4.3496	4.3344	4.3256	4.3199	4.2987

D. Correlation Coefficient

The value of correlation indicates how much the denoised signal resembles its original version.

$$\rho(\hat{s}, s) = \frac{\operatorname{cov}(\hat{s}, s)}{\sqrt{\operatorname{cov}(\hat{s}, \hat{s}) \cdot \operatorname{cov}(s, s)}} \qquad (26)$$

This criterion cancels out the influence of amplitude in ICA, the closer the correlation coefficient to 1, the better the denosing performance is.

For each SNR value, 100 independent noise sequences are generated and the correlation coefficients are calculated in Fig. 5.

From Fig. 5, performance of the proposed method remains good when signal-to-noise is low, while other methods degrade substantially. When input SNR=-5 dB, the correlation coefficient of the noisy speech and the speech signal is 0.4903, and that of the denoised result by simply EMD approaches and the speech signal becomes 0.5813. In an ascending order, the correlation coefficient reaches up to 0.6510, 0.6539, 0.6686 and 0.7237 respectively by wavelet, EMD-wavelet, EMD-ICA and EMD-wavelet-ICA approaches. The proposed method can achieve almost 47% performance enhancement compared to the noisy speech.

E. Performance Comparison between Fast ICA-R, ICA-R and Fixed-point Algorithms

Fixed-point ICA (FastICA) is the most widely used ICA algorithms in practice because of its satisfactory convergence property [22]. We make a comparison between the performance of the proposed method and other versions of it, which employ FastICA and original ICA-R in Step4 respectively. Note that we have to use prior-knowledge to select the desired IC by visual inspection when using FastICA.
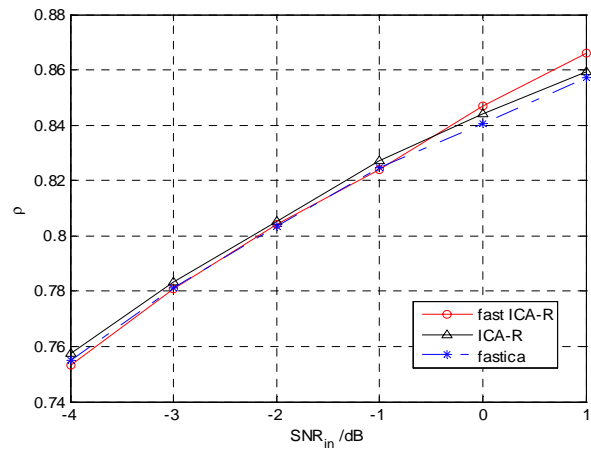
Although the application of fast ICA-R doesn't exhibit noticeable superior performance compare to FastICA, its major superiority lies in its automatic. What's more, fast ICA-R algorithm can achieve almost the same performance as the original one, actually it get even better when SNR increase above 0 dB.

F. Running Time Comparison

The running time is computed for the approaches aforementioned respectively and the averaged result of 100 times experiments is given in Table. I. The noisy speech under test has the length of 7 seconds.

In Table. I, wavelet and ICA convergent very fast, while the running time of EMD is not acceptable in practical application. Simulations results of large quantities indicate that the relationship between the running time of EMD and the length of the noisy speech is non-linear, (i.e. the longer the noisy speech is, the more time it will consume). If the noisy speech is divided into segments in advance, the running time of EMD can be considerably cut down. As an illustration, the noisy speech signal corresponding to SNR=-3 dB is processed by the modified version of EMD and the averaged running time result curve of 100 times experiments is given in the Fig. 7.

Fig. 7 shows that the running time decreases substantially when the number of segments increases, but it remains stable when number of segments surpasses 8. According to this result, the noisy speech is divided into to 8 segments in advance. Afterwards, 8 segments are processed by EMD-wavelet-ICA method respectively. At last, the processed segments are joined together to get the denoised signal.
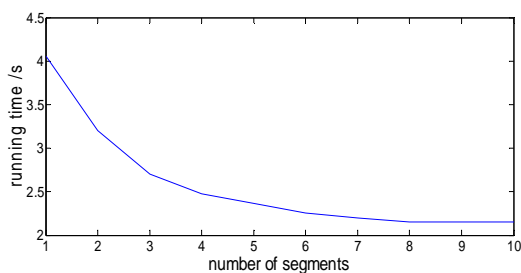
Figure 7. The relationship between number of segments with the running time of EMD
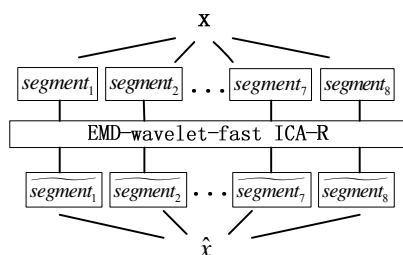

Figure 8. The framework of the proposed time-saving method

TABLE II.
THE RUNNING TIME OF THE OF THE PROPOSED TIME-SAVING METHOD

τ_p/s	SNR_{in} /dB				
	-6	-4	-2	0	2
EMD	2.1856	2.1367..	2.1435	2.1272	2.1239
EMD-wavelet-ICA-R	2.3134	2.3095	2.2747	2.2695	2.2428
EMD-wavelet-fast ICA-R	2.2583	2.2342	2.2156	2.1968	2.1932

The block diagram of the proposed time-saving method is illustrated in Fig. 8, where $\widetilde{segment_i}$ denotes the processed result of the i th segment .

The running time of this operation is shown in Table. Ⅱ, which is more acceptable now. The simulation results also demonstrate the efficiency of fast ICA-R compared with the original ICA-R. As further work, we intend to combine EMD-wavelet with fast ICA-R in a more appropriate way so as to balance the performance improvement with the computation consumption.

V. CONCLUSION

In this paper, we propose a novel approach for speech signal denoising, which merges the advantages of EMD-wavelet and fast ICA-R. Firstly, an auto-correlation criterion is proposed to determine whether the IMF is noise-dominated or not. Then wavelet thresholding is employed to process these noise-dominated IMFs and the reference signal is constructed by assembling the high-order IMFs. Finally, fast ICA-R is adopted to extract the object speech component. The total processing procedure is called EMD-wavelet-ICA.

Simulations indicate that the auto-correlation criterion we proposed is dependable. Results of the output SNR and the correlation coefficient demonstrate that EMD-wavelet-ICA exhibits superior denoising performance

compared to other approaches, especially when signal-to-noise ratio is low. In addition, constructing the reference signal of ICA-R by assembling high-order IMFs is verified to be practicable. What's more, the running time is almost twice faster compared to the existing approaches because of the efficiency of fast ICA-R and the pre-segmentation of the noisy speech.

Based on these preliminary results, we also attempt to evaluate the methods with different types of noises so as to make further improvement.

REFERENCES

[1] K. Khaldi, A. O. Boudraa, A. Alouane, A. Bouchikhi. M. T-H. Alouane, and E-H. S. Diop, "Speech signal noise reduction by EMD," International Symposium on Communications Control and Signal Processing, Mar. 2008, pp.1155-1158.

[2] N. E. Huang et al. "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," Proc. R. Soc. Lond. A, vol. 454A, pp. 903–995, Mar.1998.

[3] A. O. Boudraa, J .C .Cexus, and Z. Saidi, "EMD-based signal noise reduction," International Journal of Signal Processing, vol. 1, Jan. 2004, pp.33-37.

[4] M. A. Kabir and C. Shahnaz, "Denosing of ECG signals based on noise reduction algorithms in EMD and wavelet domains," Biomedical Signal Processing and Control, May. 2012, pp.481-489.

[5] Y. Kopsinis and S. McLaughlin. "Development of EMD-based denoising methods inspired by wavelet thresholding," IEEE Transactions on Signal Processing, vol. 57, Apr. 2009, pp. 1351-1362.

[6] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," Neural Networks, vol. 13, 2000, pp. 411-430.

[7] H. Li and Y. Sun, "Denoising by ICA based on EMD virtual Channel," Journal of Beijing University of Posts and Telecommunications, vol.30, Oct.2007, pp.33-36.

[8] M. E. Davies and C. J. James, "Source separation using single channel ICA," Signal Processing, vol. 87, Aug. 2007, pp. 1819-1832.

[9] B. Mijović, M. D. Vos, I. Gligorijević, J. Taelman, and S. V. Huffel, "Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis," IEEE Transactions on Biomedical Engineering, vol. 57, Sept. 2010, pp. 2188-2196.

[10] J. Liu, S. Yang, and S. Piao, "The single channel seismic-acoustic signal blind-source separation method based on EEMD," Journal of Harbin Engineering University, vol.32, Feb.2011, pp. 194-199.

[11] W. Lu and J. C. Rajapakse, "ICA with reference," Proc. 3rd International Conference on Independent Component Analysis and Blind Source Separation, 2001, pp.120-125.

[12] Y. Qi, G. Wang, and M. Yu, "A noise reduction method for speech signal combining ICA-R and EMD-wavelet," International Conference on Mechatronic Sciences, Electric Engineering and Computer, Dec. 2013.

[13] Z. Wu and N .E .Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," Advances in Adaptive Data Analysis 1, vol. 1, Jan.2009, pp. 1-41.

[14] M. Biswas, "An image denoising threshold estimation method," Advances in Computer Science and its Applications, vol.2, Mar. 2013, pp. 377-381.

[15] J. Li, F. Liu, H. Xu, and F. Wang, "Speech enhancement algorithm based on Hilbert-Huang and wavelet," Informatics and Management Science, Springer London, 2013.

[16] A. Hyvärinen, "Independent component analysis: recent advances," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371, vol.1984, 2013.

[17] P. Comon, "Independent component analysis: A new concept?" Signal Processing, 1994, vol. 36, pp.287-314.

[18] K. Mohanaprasad and P. Arulmozhivarman, "Comparison of Fast ICA and gradient algorithms of independent component analysis for separation of speech signals," International Journal of Engineering and Technology, 2013, pp. 3196-3202.

[19] Q. H. Lin, Y. R. Zheng, F. L. Yin, H. L. Liang, and V. D. Calhoun, "A fast algorithm for one-unit ICA-R," Information Sciences, vol.177, 2007, pp.1265-1275.

[20] Q. H. Lin, Y. R. Zheng, F. L. Yin, and H. L. Liang, "Speech segregation using constrained ICA," Advances in Neural Networks–ISNN , 2004, pp. 755-760.

[21] N. E. Huang and Z. Wu, "Statistical significance test of intrinsic mode functions," in "Hilbert-Huang Transform and Its Applications" Eds., 1st ed. Singapore, World Scientific, 2005.

[22] X. Huang and H. Zeng, "Efficient variant of FastICA algorithm for speech features extraction," Green Communication and Networks, Springer Netherlands, 2012, pp.929-935.

Yangyang Qi was born in Fuzhou, China in 1989. He received the B.S.E degree from Institute of Communication Engineering of Nanjing University of Posts and Telecommunications in June 2012.

Since September 2012, he has been working towards the M.S. degree in Institute of Communication Engineering of PLA Univ. of Sci &Tech. His current research interest is blind signal processing.

Fuqiang Yao was born in Anhui, China, in 1957. He received his M.S. degree and Ph.D. degree from Xidian University, China, in 1990 and 1993.

He is currently a researcher with Nanjing Telecommunication Technology Institute, China. He has authored two books and more than 120 journal and international conference papers. He is the holder of 13 granted patents. His current research interest is broadband communication systems.

Miao Yu was born in Jilin, China, on August 29, 1975. He received M.Sc. and Ph.D. degrees in information and communication engineering from Zhejiang University, Hangzhou, China, in 2005 and 2009 respectively.

His research interests include spread spectrum communication and blind signal processing.