

An Extended Speech De-noising Method Using GGM-Based ICA Feature Extraction

Wei Kong, Yue Zhou, and Jie Yang

Institute of Image Processing & Pattern Recognition, Shanghai Jiaotong University,
Shanghai 200030, China
{kongwei, zhoyue, jieyang}@sjtu.edu.cn

Abstract. ICA (independent component analysis) feature extraction is an efficient sparse coding method for noise reduction. In many ICA-based de-noising processing, however, they need noise-free source data to train the basis vectors as a priori knowledge. The noise-free data is always not acquirable in practice. In this paper, the generalized Gaussian model (GGM) is proposed as the p.d.f. estimator in ICA to extract the basis vectors directly from the noisy observation, since GGM can easily characterize a wide class of non-Gaussian statistical distributions. Simultaneously, the distribution of the coefficients learned by GGM is benefit for obtaining the shrinkage functions. The de-nosing experiments of noisy speech signals show that the proposed method is more efficient than conventional methods in the environment of additive white Gaussian noise. It demonstrates that the proposed method offer an efficient approach for detecting weak signals from the noise environment.

1 Introduction

Sparse coding is a method for finding a representation of data in which only a small number of the components are significantly non-zero. Because of this one may assume that the activities of components with small absolute values are noise or redundancy information and set them to zero, retaining just a few components with large activities. So such a representation is benefit for redundancy reduction and de-noising. In the work of [2]-[5], ICA has been successfully applied in extracting efficient speech features since it can extract statistical independent vectors based on the higher order statistics from data, and it was proved an efficient sparse coding method.

In many ICA-based speech de-noising works, however, the de-noising process of noisy speech signals needs noise-free source data to train the ICA basis vectors as a priori knowledge. Unfortunately, the noise-free speech signal is always not acquirable in practice. In this paper, the generalized Gaussian model (GGM) was introduced in ICA algorithm to estimate the p.d.f. of coefficients. By inferring only one parameter q , ICA algorithm can extract the efficient basis vectors directly from the noisy signals since GGM provide a general method for modeling non-Gaussian distribution. At the same time the shrinkage function can be obtained from the p.d.f. of each coefficient. Then using the maximum likelihood (ML) method on the non-Gaussian variables corrupted by additive white Gaussian noise, we show how to apply the shrinkage method on the coefficients to reduce noise. Experiments of noisy male and female speech signals show that even in the intensive noise environment, our method shows much improvement on the signal to noise ratio (SNR) after de-noising.

2 ICA Feature Extraction Using GGM

The success of the ICA learning algorithm depends highly on how well it can model the underlying statistical distribution of data. Therefore in this paper the GGM proposed by Miller J. H. et al [6] is introduced in ICA algorithm as the model for non-Gaussian coefficients. It can give different distribution models for different variables, so it is more accurate than the ICA algorithms with fixed model. Using this model in ICA, It shows that the estimation of the density of the independent coefficients is conveniently and accurately.

2.1 The Generalized Gaussian Model

The GGM models a family of density functions that is peaked and symmetric at the mean, with a varying degree of normality in the following general form [7]

$$p_g(s | \theta) = \frac{\omega(q)}{\sigma} \exp[-c(q) | \frac{s - \mu}{\sigma} |^q], \quad \theta = \{\mu, \theta, q\} \tag{1}$$

where

$$c(q) = [\frac{\Gamma[3/q]}{\Gamma[1/q]}]^{q/2} \tag{2}$$

and

$$\omega(q) = \frac{\Gamma[3/q]^{1/2}}{(2/q)\Gamma[1/q]^{3/2}} \tag{3}$$

$\mu = E[s], \sigma = \sqrt{E[(s - \mu)^2]}$ are the mean and standard deviation of the data respectively, and $\Gamma[\cdot]$ is the Gamma function. By inferring q , a wide class of statistical distributions can be characterized. The Gaussian, Laplacian, and strong Laplacian (such as speech signal) distributions can be modeled by putting $q = 2, q = 1,$ and $q < 1$ respectively. The exponent q controls the distribution's deviation from normal. When $q=2$, the distribution is the standard normal fig. 1 gives the examples of the exponential power distribution for various values of q .

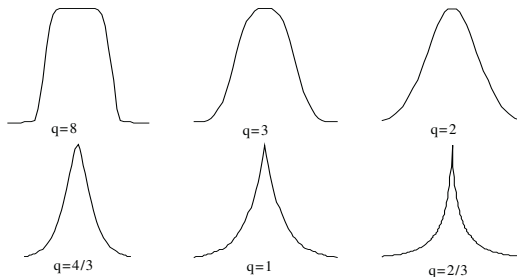


Fig. 1. Generalized Gaussian model for various exponent q

For the purposes of finding the basis functions and the parameter q in ICA, zero mean and unit variance is assumed. The problem then becomes to estimate the value

of q from the data. This can be accomplished by simply finding the maximum posteriori value q . The posterior distribution of q given the observations $\mathbf{x}=\{x_1, \dots, x_n\}$ is

$$p(q | x) \propto p(x | q)p(q) \tag{4}$$

where the data likelihood is

$$p(x | q) = \prod_n \omega(q) \exp[-c(q) | x_n |^q] \tag{5}$$

and $p(q)$ defines the prior distribution for q . Gamma function $\Gamma[\cdot]$ is used as $p(q)$ here. More details can be seen in the work of Box and Tiao (1973) [8] on inference with the exponential power distribution.

Fig. 2. shows some estimation examples of the values of q in GGM for different distributions by the maximum posterior.

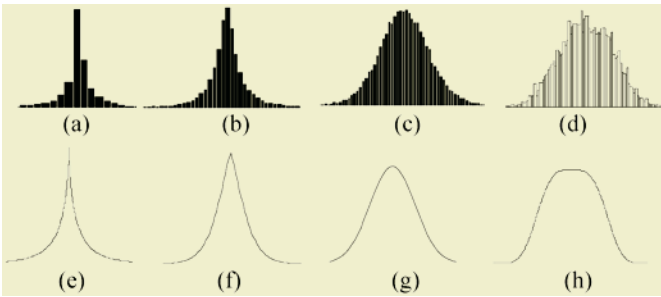


Fig. 2. (a)-(d) histogram of the distribution of s_1, s_2, s_3 and s_4 respectively, (e) $q=2/3$, (f) $q=4/3$, (g) $q=2$, (h) $q=7/2$

2.2 ICA Feature Extraction Using GGM

To extract basis functions from speech signals, ICA algorithm is applied to a number of speech segments

$$x = As = \sum_{i=1}^N a_i s_i \tag{6}$$

ICA algorithm is performed to obtain the estimation of independent components s from speech segments x by the un-mixing matrix W

$$u = Wx \tag{7}$$

where u is the estimation of independent components s . Basis functions A can be calculated from the ICA algorithm by the relation $A=W^{-1}$.

By maximizing the log likelihood of the separated signals, both the independent coefficients and the unknown basis functions can be inferred. The learning rules is represented as

$$\Delta W \propto \frac{\partial \log p(s)}{\partial W} W^T W = \eta [I - \varphi(s)s^T] W \tag{8}$$

here $W^T W$ is used to perform the natural gradient, it simplifies the learning rules and speeds convergence considerably. The vector $\varphi(s)$ is a function of the prior and is

defined by $\varphi(s) = \frac{\partial \log p(s)}{\partial s}$. In the case of the generalized Gaussian distribution (eq. 1), the vector $\varphi(s)$ can be derived as

$$\varphi_i(s_i) = -qc\sigma_i^{-q} |s_i - \mu_i|^{q-1} \text{sign}(s_i - \mu_i) \tag{9}$$

where $\mu_i = E[s_i], \sigma_i = \sqrt{E[(s_i - \mu_i)^2]}$ are the mean and standard deviation of the data respectively. Using the learning rule (8) the un-mixing matrix W is iterated by the natural gradient until convergence is achieved.

3 Speech De-noising Based on Maximum Likelihood Estimation

The GGM-based ICA algorithm in section 2.2 has been used to extract the basis vectors of noisy speech when the noise-free speech cannot be obtained, and the p.d.f. of the coefficients $p(s)$ learned by the GGM are got simultaneously. In the noise environment, denote y as the noisy coefficient of a basis vector, s as the original noise-free version of coefficient of basis vector, and v as a Gaussian noise with zero mean and variance σ^2 . Then the variable y can be describe as

$$y = s + v \tag{10}$$

We want to estimate s from the only observed noisy coefficient y . Denote p as the probability of s , and $f = -\log p$ as its negative log-density, the estimator of s can be obtained by the maximum likelihood (ML) method

$$\hat{s} = \arg \min_s \frac{1}{2\sigma^2} (y - s)^2 + f(s) \tag{11}$$

Assuming $f(\cdot)$ to be strictly convex and differentiable, the ML estimation gives the equation

$$\hat{s} = h(y) \tag{12}$$

where the nonlinear function $h(\cdot)$ is called as *shrinkage* function, and the inverse is given by

$$h^{-1}(s) = s + \sigma^2 f'(s) \tag{13}$$

Thus, the estimation of s is obtained by inverting a certain function involving $f'(\cdot)$. Since $f(\cdot)$ is a function of p , the probability of s has been obtained by GGM in ICA feature extraction, the shrinkage function can be obtained easily.

To recover the de-noised speech signal from the noisy source three steps are needed. Firstly, By using GGM-based ICA, we can obtain the un-mixing matrix W and the p.d.f. of the corresponding coefficients $p(s)$ at the same time. From the experiments, it shows that the coefficients of the basis vectors extracted from noisy speech have sparse distributions. Secondly, the shrinkage functions can be estimated by $p(s)$ by eq. (13), and the de-noised coefficients can be calculated by $\hat{s} = h(y)$. Finally, recover the de-noised speech signal by $\hat{x} = W^{-1} \hat{s} = A \hat{s}$.

This method is closed related to the wavelet shrinkage method. However, the sparse coding based on ICA may be viewed as a way for determining the basis and

corresponding shrinkage functions base on the data themselves. Our method use the transformation based on the statistical properties of the data, whereas the wavelet shrinkage method chooses a predetermined wavelet transform. And the second difference is that we estimate the shrinkage nonlinearities by the ML estimation, again adapting to the data themselves, whereas the wavelet shrinkage method use fixed threshold derived by the mini-max principle.

4 Experiments

Noisy male speech and noisy female speech signals mixed with white Gaussian noise were used to test the performance of the proposed method. The sampling rates are both 8kHz and 20000 samples of each noisy speech signal is used. The first step is the feature extraction of the noisy signals using the GGM-based ICA algorithm described in section 2. For each noisy speech signal, the mean was subtracted (eq.14) and then 500 vectors of length 40 (5ms) were generated, and each segment was pre-whitened to improve the convergence speed (eq.15).

$$x = x - E\{x\} \quad (14)$$

$$v = E\{x x^T\}^{-1/2} x \quad (15)$$

This pre-processing removes both first- and second-order statistics from the input data, and makes the covariance matrix of x equal to the identity matrix, where x denoted as the observed noisy signals. The adaptation of the un-mixing matrix W started from the 40×40 identity matrix and trained through the 500 vectors. The learning rate was gradually decreased from 0.2 to 0.05 during the iteration. Then the un-mixing matrix W was extracted by the learning rule (8), and it was used as the filter in the de-noising processing. To judge the results of the de-noising, the signal-to-noise ratio (SNR) is used

$$SNR_i = 10 \log \left| \frac{\sum_{t=1}^N Signal(t)^2}{\sum_{t=1}^N Noise(t)^2} \right| \quad (16)$$

In the first experiment, the noisy male speech signals corrupted by four different intensity of additive white Gaussian noise were used to test the de-noising method. Their SNR of the input noisy signals are 4.8004, 0.4854, -5.0242 and -12.6541dB respectively. The output SNR results of the de-noised speech signals are 8.0815, 6.0971, 3.4017 and -0.2596dB respectively, it can be seen that the SNR have much improvement. Fig. 3 shows the de-noising results of the noisy male speech with the input SNR of -5.0242dB and it was compared to the median filter and the wavelet filter method.

The de-noising results of these four noisy male speech signals with different intensity Gaussian noise are shown in table 1. Where SNR_{in} denotes the input SNR of the noisy male speech signals and SNR_{out} denotes the output SNR of the de-noised male speech signals.

And the de-noising results of these four noisy female speech signals with different intensity Gaussian noise are shown in table 2.

Fig. 5 shows the SNRs in different intensities of the additive white Gaussian noise for male speech experiments and female speech experiments respectively. It shows that our method of extracting basis vectors directly from the noisy signals based on GGM is efficient and always better than conventional methods.

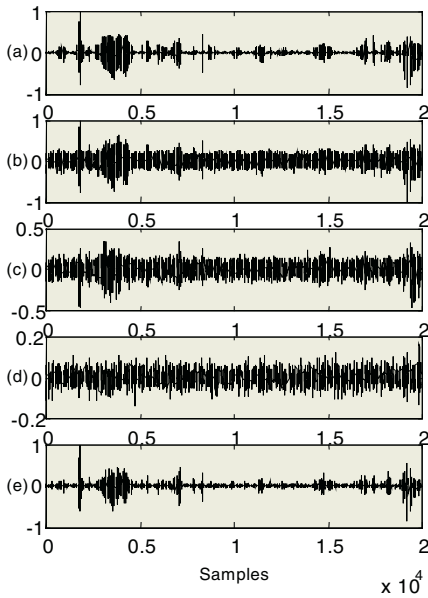


Fig. 3. (a) noisy-free male speech signals, (b) noisy male speech signals with input SNR of -5.0242dB , (c) de-noising result of median filter with $n=3$, output SNR is -4.8481dB , (d) de-noising result of wavelet filter with db3 and $n=3$, output SNR of this method is -1.8335dB , (e) de-noising result of our method, output SNR is 3.4017 dB

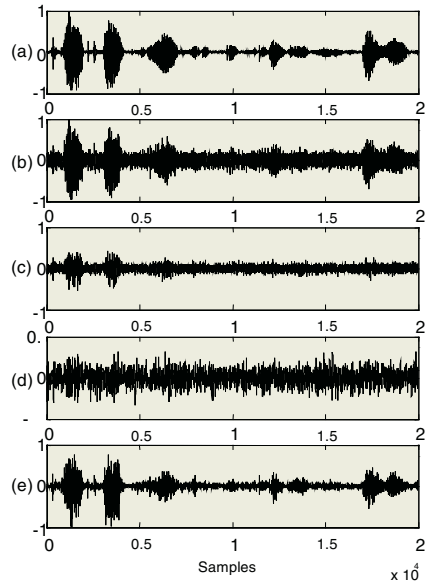


Fig. 4. (a)noisy-free female speech signals, (b)noisy female speech signals with input SNR of -7.0775dB , (c) de-noising result of median filter with $n=3$, output SNR is -4.6386dB , (d) de-noising result of wavelet filter with db3 and $n=3$, output SNR of this method is -2.2646dB , (e) de-noising result of our method, output SNR is 4.7150 dB

Table 1. Signal-to-noise ratio (SNR) of the de-noised male speech signals

SNR _{in} of noisy male speech (dB)	SNR _{out} of median filter (dB)		SNR _{out} of wavelet filter,db3,n=3 (dB)	SNR _{out} of our method (dB)
	n=3	n=5		
4.8004	-1.7077	-2.7722	-0.6592	8.0815
0.4854	-2.8087	-3.4863	-1.0537	6.0971
-5.0242	-4.8481	-4.8504	-1.8335	3.4017
-12.6541	-8.9528	-7.7528	-3.6795	-0.2596

Table 2. Signal-to-noise ratio (SNR) of the de-noised female speech signals

SNR _{in} of noisy female speech (dB)	SNR _{out} of median filter (dB)		SNR _{out} of wavelet filter,db3,n=3 (dB)	SNR _{out} of our method (dB)
	n=3	n=5		
2.6087	-0.0310	-4.4573	-0.8806	6.5623
-1.6538	-1.7842	-5.1106	-1.3522	5.8288
-7.0775	-4.6386	-6.3778	-2.2646	4.7150
-14.5406	-9.6556	-9.2051	-4.3425	0.7533

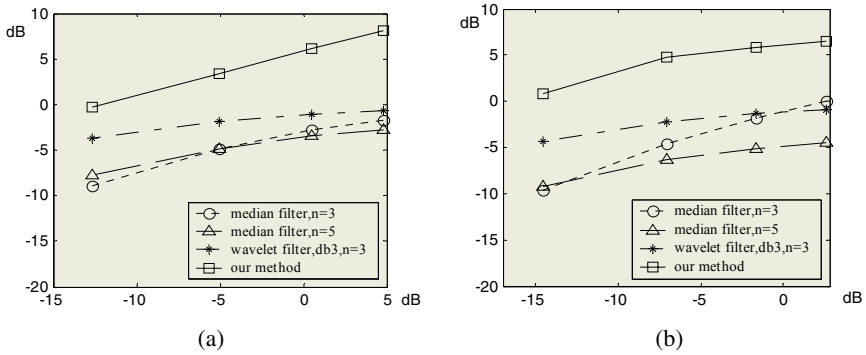


Fig. 5. (a) the SNR comparison of several de-noising method on noisy male speech signals, (b) the SNR comparison of several de-noising method on noisy female speech signals

5 Conclusions

How to extract efficient basis vectors directly from the observed noisy speech signals is the key objective of noisy speech de-noising since the corresponding noise-free signal is always not acquirable in practice. So in this paper, ICA feature extraction using generalized Gaussian model (GGM) is proposed to extract basis vectors directly from noisy data. By inferring only one parameter, different kinds of non-Gaussian distributions of coefficients can be characterized. Sparse coding is achieved by ICA feature extraction. And the ICA features and the shrinkage functions can be obtained at the same time. By shrinkage the absolute values of the sparse components to wards zero, noise can be reduced. Experiments on noisy male and female speech signals show that the proposed method can efficiently remove the additive white Gaussian noise.

References

1. A. Hyvärinen, Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation. Technical Report A51, Helsinki University of Technology, Laboratory of Computer and Information Science, 1998
2. Te-Won Lee, Gil-Jin Jang, The Statistical Structures of Male and Female Speech Signals, in Proc. ICASSP, (Salt Lake City, Utah), May 2001
3. Jong-Hawn Lee, Ho-Young Jung, Speech Feature Extraction Using Independent Component Analysis, in Proc. ICASP, Istanbul, Turkey, June, 2000, Vol. 3, pp: 1631-1634
4. Anthony J Bell, Terrence J Sejnowski, Learning the Higher-order structure of a nature sound, Network: Computation in Neural System 7 (1996), 261-266
5. Gil-Jin Jang, Te-won Lee, Learning statistically efficient features for speaker recognition, Neurocomputing, 49 (2002): 329-348
6. Miller J. H. & Thomas J. B., Detectors for Discrete-Time Signals in Non-Gaussian noise, IEEE Transactions on Information Theory, Vol IT-18, no. 2, March 1972. Page(s) 241-250
7. Te-Won Lee, Michael S. Lewicki, The Generalized Gaussian Mixture Model Using ICA, in international workshop on Independent Component Analysis (ICA'00), Helsinki, Finland, June 2000, pp: 239-244