Efficient Feature Extraction and De-noising Method for Chinese Speech Signals Using GGM-Based ICA

Yang Bin and Kong Wei

Information Engineering College, Shanghai Maritime University, Shanghai 200135, China
binyang@cie.shmtu.edu.cn, kongwei@sjtu.edu.cn

Abstract. In this paper we study the ICA feature extraction method for Chinese speech signals. The generalized Gaussian model (GGM) is introduced as the p.d.f. estimator in ICA since it can provide a general method for modeling non-Gaussian statistical structure of univariate distributions. It is demonstrated that the ICA features of Chinese speech are localized in both time and frequency domain and the resulting coefficients are statistically independent and sparse. The GGM-based ICA method is also used in extracting the basis vectors directly from the noisy observation, which is an efficient method for noise reduction when priori knowledge of source data is not acquirable. The de-nosing experiments show that the proposed method is more efficient than conventional methods in the environment of additive white Gaussian noise.

1 Introduction

Chinese is a typical tonal and syllabic language, in which each Chinese character corresponds to a monosyllable and basically has a phoneme structure with a lexical tone. Each Chinese character has four lexical tones (Tone1, Tone2, Tone 3, and Tone 4) and a neutral tone. There are about 400 toneless Chinese syllables and about 1,300 toned Chinese syllables. How to extract efficient features from Chinese speech signals is a key task of Chinese speech coding, de-noising and recognition.

Nowadays, many efforts have gone into finding learning algorithms to obtain the statistical characteristics of speech and sound signals. However, these commonly used features have the limitations that they are sensitive only to second-order statistics since they all use correlation-based learning rules like principal component analysis (PCA). The failure of correlation-based learning algorithm is that they are typically global and reflect only the amplitude spectrum of the signal and ignore the phase spectrum. The most informative features of sound signals, however, require higher-order statistics for their characterization [1-4]. For this reason, we study the ICA feature extraction method on Chinese speech signals in this paper. The generalized Gaussian model was introduced here to provide a general method for modeling non-Gaussian statistical structure of the resulting coefficients which have the form of $p(x) \propto \exp(-|x|^q)$. By inferring q, a wide class of statistical distributions can be characterized. By comparing the ICA basis with DFT, DCT and PCA basis, it can be seen that the proposed method is more efficient than conventional features.

M. Lazo and A. Sanfeliu (Eds.): CIARP 2005, LNCS 3773, pp. 925 – 932, 2005.

The advantage of GGM-based ICA method is also applied in the de-noising of Chinese speech signals even when the trained priori knowledge of source data is not acquirable. Not only the ICA features but also the de-noising shrinkage function can be obtained from the GGM-based ICA sparse coding. Using the maximum likelihood (ML) method on the non-Gaussian variables corrupted by additive white Gaussian noise, we show how to apply the GGM-based shrinkage method on the coefficients to reduce noise. Experiment of noisy male Chinese speech signals shows that our de-noising method is successful in improving the signal to noise ratio (SNR).

2 ICA Feature Extraction Using GGM

In ICA feature extraction methods, the source speech signal is represented as segments

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

Where $\boldsymbol{A}$ is defined as 'basis vector' of source signals, and $\boldsymbol{s}$ is its corresponding coefficient. ICA algorithm is performed to obtain the estimation of independent components $\boldsymbol{s}$ from speech segments $\boldsymbol{x}$ by the un-mixing matrix $\boldsymbol{W}$

$$\boldsymbol{u} = \boldsymbol{W}\boldsymbol{x} \tag{2}$$

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

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

$$\Delta W \propto \frac{\partial \log p(s)}{\partial W} W^T W = \eta[I - \varphi(s)s^T]W \tag{3}$$

here $\boldsymbol{W}^T\boldsymbol{W}$ 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 $\varphi(s) = \frac{\partial \log p(s)}{\partial s}$, and $p(s)$ is the p.d.f. of s. Here we use the GGM as the p.d.f. estimator. 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 [5]

$$p_g(s \mid \theta) = \frac{\omega(q)}{\sigma} \exp[-c(q) \mid \frac{s-\mu}{\sigma} \mid^q], \qquad \theta = \{\mu, \theta, q\} \tag{4}$$

where

$$c(q) = [\frac{\Gamma[3/q]}{\Gamma[1/q]}]^{q/2} \tag{5}$$

and

$$\omega(q) = \frac{\Gamma[3/q]^{\frac{1}{2}}}{(2/q)\Gamma[1/q]^{\frac{3}{2}}} \tag{6}$$

$\mu = E[s], \sigma = \sqrt{E[(s-\mu)^2]}$ 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.

For the purposes of finding the basis functions, 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 $\boldsymbol{x}=\{x_1,\ldots,x_n\}$ is

$$p(q \mid x) \propto p(x \mid q)\, p(q) \tag{7}$$

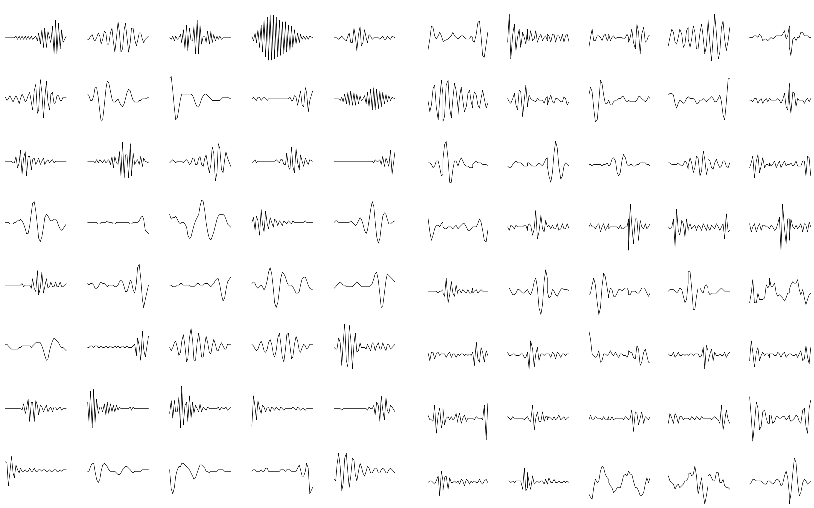
where the data likelihood is

$$p(x \mid q) = \prod_n \omega(q) \exp[-c(q) \mid x_n \mid^q] \tag{8}$$

and $p(q)$ defines the prior distribution for q, here Gamma function $\Gamma[\cdot]$ is used as $p(q)$.

In the case of the GGM, the vector $\varphi(s)$ in eq.3 can be derived as

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

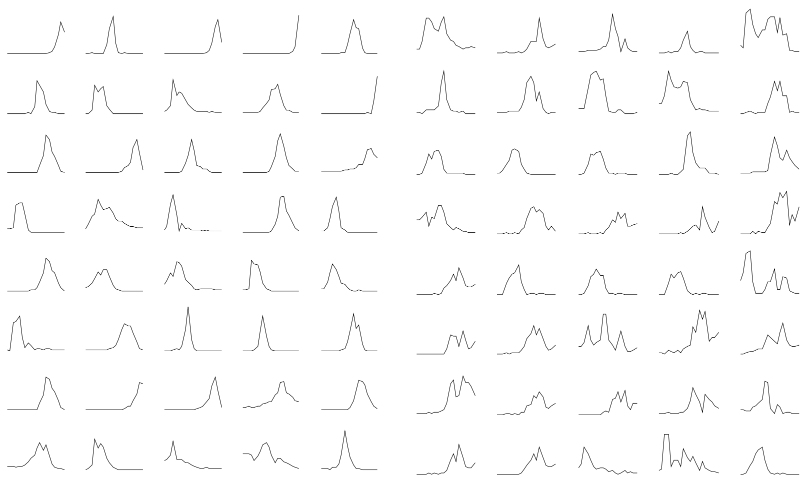


(a) (b)

Fig. 1. (a)-(b) Some basis vectors of male and female Chinese speech signals

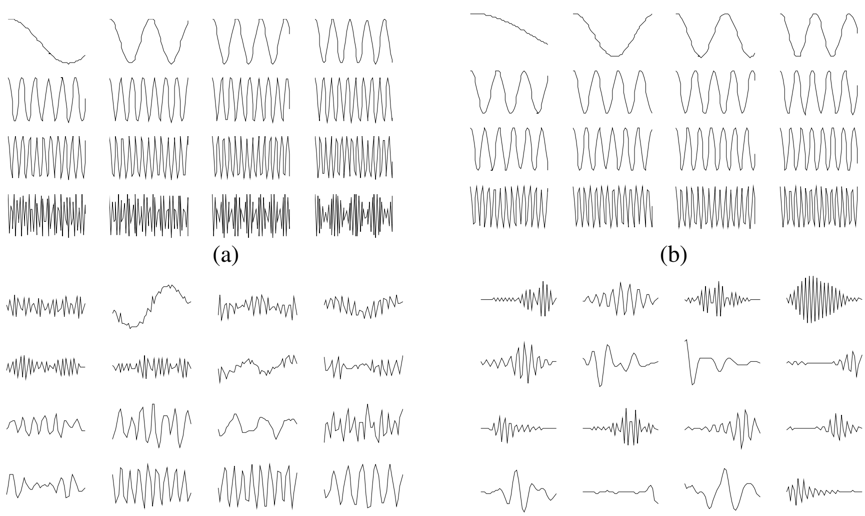
Using the learning rule eq. 3 the un-mixing matrix $\boldsymbol{W}$ is iterated by the natural gradient until convergence is achieved.

To learn the basis vector, one male Chinese speech signals and one female Chinese speech signals were used. The sampling rates of the original data are both 8kHz. Fig.1 (a) and (b) show some of the basis vector of the male and female Chinese speech



(a) (b)

Fig. 2. (a)-(b) The frequency spectrum of fig.1 (a) and (b)



(a) (b) (c) (d)

Fig. 3. Comparison of DFT, DCT, PCA and ICA basis vector of male Chinese speech signal, (a) DFT basis vector, (b) DCT basis vector, (c) PCA basis vector, (d) ICA basis vector

signals learned by the GGM-based ICA method. Fig.2 shows the frequency spectrum of fig.1 (a) and (b) respectively. It can be seen that the ICA basis vectors of Chinese speech signals are localized both in time and frequency domain.

For comparison, discrete Fourier transform (DFT), discrete cosine transform (DCT), and principal component analysis (PCA) basis vectors as conventional methods are adopted. Fig.3 compares the waveforms of the DFT, DCT and PCA basis with the ICA basis.16 basis functions of male Chinese speech signals for each method are displayed.

From fig. 3 (a)-(d) we can see that the DFT and DCT basis look similar and they are spread all over the time axis. For different signals the DFT and DCT basis are fixed. PCA basis is data driven and exhibits less regularity and global. However, the ICA basis functions are localized in time and frequency, thus they reflect both the phase and frequency information inherent in the data.

3 Speech De-noising Using GGM-Based ICA

ICA feature extraction is wildly used in de-noising of image and speech signals since ICA is an efficient sparse coding method for finding a representation of data [6, 7]. In these methods, however, the trained basis vectors were needed and applied for the removal of Gaussian noise. 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

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

Denote *p* as the probability of *s*, and $f = -\log p$ as its negative log-density, we want to estimate *s* from the observed noisy coefficient *y*. 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 $f(\bullet)$ to be strictly convex and differentiable, the ML estimation gives the equation

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

where the nonlinear function h(•) 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 $f'(\bullet)$. Since $f(\bullet)$ is a function of *p*.

There are two difficulties in this method. One is: the noise-free source data is needed to train the ICA basis vectors as a priori knowledge. Unfortunately, the corresponding noise-free signals are always not acquirable in practice. The other is how to efficiently estimate the p.d.f. of the coefficient vector s which is the key of estimating $\hat{s}$. To solve these two problems the GGM-based ICA algorithm in section 2 is used to extract the basis vectors directly from noisy speech signals when the noise-free signals cannot be obtained. It is fortunately that the p.d.f. of the coefficients $p(s)$ can be learned by the GGM simultaneously since the parameter q of the GGM is determined during the ICA feature extraction.

To recover the de-noised speech signal from the noisy source three steps are needed. Firstly, extract the ICA basis vector directly from the noisy speech signals by using GGM-based ICA. The p.d.f. of the corresponding coefficients $p(s)$ are obtained at the same time. It is demonstrated that the coefficients of the basis vectors extracted directly 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 Chinese speech signals mixed with white Gaussian noise were applied to perform the proposed method. The sampling rate is 8kHz and 64000 samples are used. The first step is the feature extraction of the noisy signals using the GGM-based ICA algorithm described in section 2. For the noisy speech signal, the mean was subtracted (eq.14) and then 1000 vectors of length 64 (8ms) were generated, and each segment was pre-whitened to improve the convergence speed (eq.15).

$$\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{E}\{\boldsymbol{x}\} \tag{14}$$

$$\boldsymbol{v} = \boldsymbol{E}\{\boldsymbol{x}\,\boldsymbol{x}^{\boldsymbol{T}}\}^{-1/2}\boldsymbol{x} \tag{15}$$

This pre-processing removes both first- and second-order statistics from the input data, and makes the covariance matrix of $\boldsymbol{x}$ equal to the identity matrix, where $\boldsymbol{x}$ denoted as the observed noisy signals. The adaptation of the un-mixing matrix $\boldsymbol{W}$ started from the 64×64 identity matrix and trained through the 1000 vectors. The learning rate was gradually decreased from 0.2 to 0.05 during the iteration. The signal-to-noise ratio (SNR) is used to judge the results of the de-noising

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

Fig. 4 shows the noisy male Chinese speech signals with the input SNR of 6.3850dB and the de-noising results of wavelet method (db3, n=3) and our proposed method in (b), (c) and (d) respectively. For comparison, the corresponding noise-free signal is given by (a). The SNR of the input noisy signal is 6.3850. The output SNR of wavelet and GGM-based ICA method are 10.5446 and 12.9910 respectively. It can be seen that the de-noising result of the proposed method is better than that of wavelet de-noising method.

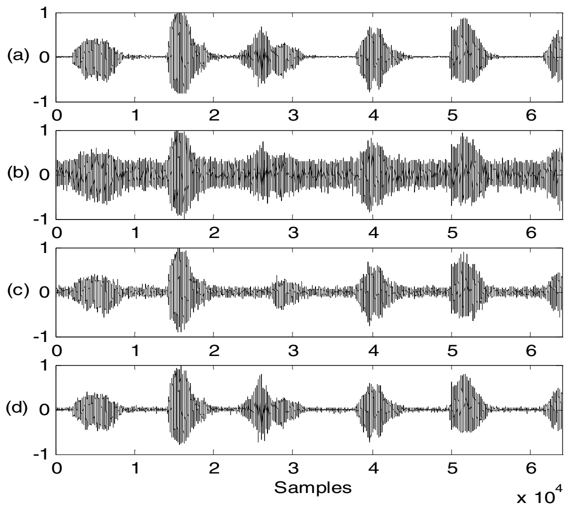


Fig. 4. The de-noising results of male Chinese speech signals, (a) noise-free male Chinese speech signal, (b) noisy male Chinese speech signal, (c) the de-noising result of wavelet, (d) the de-noising result of GGM-based ICA

5 Conclusions

In this paper, we obtained efficient feature extraction method for Chinese speech signals. It is demonstrated that the GGM-based ICA features are localized both in time and frequency domain. This efficient ICA feature extraction method was also applied to the de-noising of Chinese speech signals and demonstrated better performance than wavelet de-noising method. The proposed de-noising method can be directly used in practice since it does not need the noise-free signals to train the priori knowledge. The experiment on noisy male Chinese speech signal shows that the proposed method is efficient to remove the additive white Gaussian noise.

References

1. Te-Won Lee, Gil-Jin Jang, The Statistical Structures of Male and Female Speech Signals, in Proc. ICASSP, (Salt Lack City, Utah), May 2001
2. 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

3. Anthony J Bell, Terrence J Sejnowski, Learning the Higher-order structure of a nature sound, Network: Computation in Neural System 7 (1996), 261-266
4. Gil-Jin Jang, Te-won Lee, Learning statistically efficient features for speaker recognition, Neurocomputing, 49 (2002): 329-348
5. 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
6. 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
7. Hyvärinen A., Hoyer P., Oja E., Sparse code shrinkage: Denoising by nonlinear maximum likelihood estimation, Advances in Neural Information Processing System 11 (NIPS'98), 1999