Speech Signal Enhancement Based on MAP Algorithm in the ICA Space

Xin Zou, Peter Jančovič, *Member, IEEE*, Ju Liu, *Member, IEEE*, and Münevver Köküer, *Member, IEEE*

***Abstract*—This paper presents a novel maximum *a posteriori* (MAP) denoising algorithm based on the independent component analysis (ICA). We demonstrate that the employment of individual ICA transformations for signal and noise can provide the best estimate within the linear framework. The signal enhancement problem is categorized based on the distribution of signal and noise being Gaussian or non-Gaussian and the estimation rule is derived for each of the categories. Our theoretical analysis shows that under the assumption of a Gaussian noise the proposed algorithm leads to some well-known enhancement techniques, i.e., Wiener filter and sparse code shrinkage. The analysis of the denoising capability shows that the proposed algorithm is most efficient for non-Gaussian signals corrupted by a non-Gaussian noise. We employed the generalized Gaussian model (GGM) to model the distributions of speech and noise. Experimental evaluation is performed in terms of signal-to-noise ratio (SNR) and spectral distortion measure. Experimental results show that the proposed algorithms achieve significant improvement on the enhancement performance in both Gaussian and non-Gaussian noise.**

***Index Terms*—Generalized Gaussian model, independent component analysis (ICA), maximum *a posteriori* (MAP) estimation, non-Gaussian noise, sparse code shrinkage, speech enhancement, Wiener filter.**

I. Introduction

THE objective of speech enhancement algorithms is to improve the quality of noise-corrupted speech signals by removing the corrupting noise. Over past several decades much research has been focused on this area.

There have been many techniques proposed for enhancement of speech signal corrupted by an additive noise. Usually, it is assumed that the noise is independent of speech signal. Spectral subtraction [1], which performs subtraction of a noise spectral estimate from a noisy speech spectrum, was suggested in early years, but has still been popular due to its computational efficiency [2], [3]. Wiener filtering, e.g., [4], [5], performs filtering of noisy speech signal by using a filter derived based on the minimum mean-square error criterion. It assumes that both the speech and noise signals have a Gaussian distribution. In recent studies, the use of maximum *a posteriori* (MAP) algorithm has been proposed, e.g., [6]–[12]. The estimation is usually carried out in a linear transformation domain. Studies were presented on the use of DCT and KLT [11], wavelet transform [9], or ICA transform [6]–[8], [10]. In the MAP-based algorithms, the distribution of speech signals is typically assumed to be non-Gaussian. Non-Gaussian model leads to a nonlinear estimation rule, and this has been reported to obtain a better performance in many cases [13]. The Laplace distribution was employed in [10] and [11]. The Gaussian mixture model was used in [9]. In [6]–[8], signals were modeled by a combination of a mildly super-Gaussian, Laplacian, and strongly super-Gaussian distributions, and the obtained algorithm was referred to as sparse code shrinkage (SCS). The MAP algorithm proposed in [10] considers a possible correlation among the ICA-domain noise components. The enhancement algorithms proposed in the above papers are based on an assumption that the noise is Gaussian distributed. An effective dealing with a non-Gaussian noise corruption as well as finding an efficient linear transformation for specific noisy conditions and a more flexible distribution modeling are still open topics that need further investigation.

In this paper, we attempt to address some of the above questions. We propose a novel MAP single-channel signal denoising algorithm that uses the ICA transformation. We demonstrate that within the linear framework the employment of individual ICA transformations for signal and noise can provide the best signal estimate. The signal enhancement problem is categorized into three situations: a) Gaussian signal and Gaussian noise; b) non-Gaussian signal and Gaussian noise; and c) non-Gaussian signal and non-Gaussian noise. The estimation rule is derived for each of the categories and the denoising capability of those situations is theoretically analyzed and compared. We also propose an employment of generalized Gaussian model (GGM), originally introduced in [14], as a flexible model for modeling a wide range of non-Gaussian distributions. The proposed algorithms are evaluated on speech signal corrupted by both Gaussian and non-Gaussian noises. The performance is presented in terms of signal-to-noise ratio (SNR) and spectral distortion measure. Experimental results show that the proposed algorithms achieve, in correspondence with our theoretical analysis, significant improvement on the enhancement performance in both Gaussian and non-Gaussian noises.

The paper is organized as follows. The general MAP algorithm based on ICA is presented in Section II. Some of the special cases of MAP algorithm are discussed in Section III. The denoising capability of the proposed algorithm is analyzed in Section IV. The GGM model and its parameter estimation is introduced in Section V. The summary of proposed algorithm is given in Section VI. Section VII presents experimental evaluation and results, and Section VIII gives conclusions.

Manuscript received February 4, 2007; revised August 29, 2007. This work was supported by EPSRC Grants EP/D033659/1 and EP/F036132/1, U.K., and by the Program for New Century Excellent Talents in University (NCET-05-0582) and Specialized Research Fund for the Doctoral Program of Higher Education (Grant 20050422017), China. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Antonio Napolitano.

X. Zou, P. Jančovič, and M. Köküer are with the Electronic, Electrical and Computer Engineering (School of Engineering), University of Birmingham, B15 2TT, Birmingham, U.K. (e-mail: xxz391@bham.ac.uk; p.jancovic@bham.ac.uk; m.kokuer@bham.ac.uk).

J. Liu is with the School of Information Science and Engineering, Shandong University, Shandong, Jinan 250100, China (e-mail: juliu@sdu.edu.cn).

Digital Object Identifier 10.1109/TSP.2007.910555

II. Maximum *A Posteriori* Estimation by ICA

Let us consider scalar random variables. Denote by x the original clean signal, and by v some additive noise. Assume that we have observed the random variable y which is the noisy version of signal x

$$y = x + v. \tag{1}$$

We are interested in estimating x from the observed noisy signal y by means of $\hat{x} = g(y)$. The MAP algorithm has been widely applied to solve this problem. Compared with the maximum-likelihood (ML) estimation algorithm, which maximizes the conditional density, the maximization in the MAP algorithm takes into account both the conditional density and the *a priori* distribution of the signal and as such may lead to better denoising performance. Using log probability density to replace the normal density function, the estimate $\hat{x}$ can be obtained by MAP estimation algorithm as

$$\hat{x} \leftarrow \arg\max_x (\ln p(y \,|\, x) + \ln p(x)) \tag{2}$$

where $p(y \,|\, x)$ is the conditional density of the observation y given x, which is the density of noise evaluated at $y - x$, i.e., $p(y \,|\, x) = p_v(y-x)$, and $p(x)$ is the *a priori* distribution density of original clean signal x, which for clarity of the following derivation we denote by $p_x(x)$.

Let us consider the entire single-channel noisy speech signal $y(t)$ is split into short-term segments, each consisting of N samples. This gives a set of N-dimensional observation vectors $\mathbf{y} = \{\mathbf{y}(1), \ldots, \mathbf{y}(T)\}$, where $\mathbf{y}(t) = [y(t), \ldots, y(t-N+1)]^T$ and $(\cdot)^T$ denotes the transpose operation. Corresponding notation is used for clean speech signal. According to (2), the estimation of $\hat{\mathbf{x}} = \{\hat{\mathbf{x}}(1), \ldots, \hat{\mathbf{x}}(T)\}$ for large $T \to \infty$ can be obtained by

$$\begin{aligned} &\{\hat{\mathbf{x}}(1), \ldots, \hat{\mathbf{x}}(T)\} \\ &\leftarrow \arg\max_{\hat{\mathbf{x}}} \left(\sum_t \ln p_v(\mathbf{y}(t) - \hat{\mathbf{x}}(t)) + \sum_t \ln p_x(\hat{\mathbf{x}}(t)) \right) \\ &\Leftrightarrow \arg\max_{\hat{\mathbf{x}}} (E\{\ln p_v(\mathbf{y}(t) - \hat{\mathbf{x}}(t))\} + E\{\ln p_x(\hat{\mathbf{x}}(t))\}). \end{aligned} \tag{3}$$

In the following derivations, for convenience, the time index t will be omitted and without causing any confusion, we just use $\mathbf{x}$ to replace the estimate $\hat{\mathbf{x}}$.

Let us first consider the term $E\{\ln p_x(\mathbf{x})\}$, some simple manipulations will yield [15]

$$E\{\ln p_x(\mathbf{x})\} = -H(p_{x,A^*}(\mathbf{x})) - \mathrm{KL}(p_{x,A^*}(\mathbf{x}) \| p_{x,A}(\mathbf{x})) \tag{4}$$

where $\mathrm{KL}(\cdot \,\|\, \cdot)$ and $H(\cdot)$ denote Kullback–Leibler divergence and entropy, respectively, the matrices A^* and A denote the mixing matrix and its estimate, respectively, and the $p_{x,A^*}(\mathbf{x})$ is the true distribution of $\mathbf{x}$, which can be expressed as

$$p_{x,A^*}(\mathbf{x}) = p_x(A^* \mathbf{s}_x) \tag{5}$$

where $\mathbf{s}_x$ are the underlying independent components of $\mathbf{x}$. For any given signal $\mathbf{x}$, the A^* and $\mathbf{s}_x$ are unique and therefore the entropy term in (4) can be considered as constant. As such, the maximization of $E\{\ln p_x(\mathbf{x})\}$ can be achieved by the minimization of $\mathrm{KL}(p_{x,A^*}(\mathbf{x}) \| p_{x,A}(\mathbf{x}))$, which would result in zero, as the KL divergence is nonnegative, if and only if the estimate A equals to A^*. It has been shown, e.g., [15], that the use of ICA principle can provide an efficient way of estimation of the underlying independent sources $\mathbf{s}$ and the corresponding mixing matrix A^* from given signals. As such the maximization of $E\{\ln p_x(\mathbf{x})\}$ can be achieved within the ICA framework. Then, by using the relationship between the density of a random variable $\mathbf{b}$ and its linearly transformed version $\mathbf{d} = W\mathbf{b}$, i.e., $p_b(\mathbf{b}) = p_d(W\mathbf{b})|\det(W)|$, the maximum log likelihood $E\{\ln p_x(\mathbf{x})\}$ can be expressed as

$$E\{\ln p_x(\mathbf{x})\}_{\max} = E\{\ln p_{s_x}(W^x \mathbf{x})\} + \ln|\det(W^x)|. \tag{6}$$

By applying the analysis as presented above, the same conclusion can be drawn for the term $E\{\ln p_v(\mathbf{y} - \mathbf{x})\}$, i.e.,

$$E\{\ln p_v(\mathbf{y}-\mathbf{x})\}_{\max} = E\{\ln p_{s_v}(W^v(\mathbf{y}-\mathbf{x}))\} + \ln|\det(W^v)|. \tag{7}$$

In (6) and (7), the W^x and W^v are ICA unmixing matrices (i.e., inverse of mixing matrices) of clean signal and noise, respectively, and p_{s_x} and p_{s_v} are the density functions of the underlying independent components of clean signal x and noise v, respectively.

Based on this analysis, we propose to perform the MAP estimation in (3) by employing two individual ICA transformations, one obtained from the signal and the other from noise. Considering a single frame estimation, the expectation operator can be removed. The ICA based MAP algorithm can be expressed in the form of

$$\begin{aligned} \mathbf{x} \leftarrow \arg\max_{\mathbf{x}} (&\ln p_{s_v}(W^v(\mathbf{y} - \mathbf{x})) + \ln p_{s_x}(W^x \mathbf{x}) \\ &+ \ln|\det(W^x)||\det(W^v)|). \end{aligned} \tag{8}$$

Since these matrices are, for a given noise and clean signal, fixed, the determinant term in (8) can be omitted. Let us denote the right term of (8) by L. The estimate of the nth sample of the signal x_n can be obtained by using a gradient method

$$\begin{aligned} x_n &\leftarrow x_n + \lambda \frac{\partial L}{\partial x_n} \\ &= x_n + \lambda \sum_{k=1}^{N} [f_v'(W_{k\cdot}^v(\mathbf{y} - \mathbf{x}))w_{kn}^v - f_x'(W_{k\cdot}^x \mathbf{x})w_{kn}^x] \end{aligned} \tag{9}$$

where λ is the step size, $f'(s) = -\partial \ln p(s)/\partial s, w_{kn} = W(k,n)$, and $W_{k\cdot} = W(k,:)$.

Equation (9) is the general case of the proposed ICA-based MAP signal enhancement algorithm and it is referred to as ICAMAP(gen) hereafter. Note that the ICAMAP(gen) algorithm derived above does not make any assumptions about the properties of the noise (i.e., Gaussian, non-Gaussian)—as such it can be used to deal with both non-Gaussian and Gaussian noises. In the case of both signal and noise being non-Gaussian, the best estimate can be obtained by using two ICA transformation matrices as shown above. When the noise is Gaussian, a single matrix can be used in place of these two matrices which can then lead to a simplified solution to the estimation in (9)—this is discussed in detail in the following section.

III. Special Cases of MAP Estimator in ICA Space

This section analyzes the denoising algorithm proposed in the previous section in two special cases: a) assuming both the signal and noise to be Gaussian distributed, and b) assuming only the noise to be Gaussian. It is shown below that the ICAMAP(gen) leads under these considerations to some well known simpler solutions.

A. Gaussian Distributed Signal and Noise

Let us consider both the signal and noise to be zero mean Gaussian distributed, i.e.,

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \tag{10}$$

where σ^2 denotes the variance. As it has been demonstrated in [16], for Gaussian distribution, the maximization in (3) is not affected by a linear transformation. As such, the estimation can be performed in the original space (i.e., the transformation matrices W^x and W^v equal to identity). In view of (8) and (10), the estimation rule for each dimension will be

$$\begin{aligned} x_n &\leftarrow \arg\max_x (\ln p_v(y_n - x_n) + \ln p_x(x_n)) \\ &\Leftrightarrow \arg\min_x \left(\frac{1}{2\sigma_v^2}(y_n - x_n)^2 + \frac{1}{2\sigma_x^2} x_n^2 \right). \end{aligned} \tag{11}$$

This minimization is equivalent to solving the following equation, i.e., setting the derivative of (11) with respect to x_n to zero

$$\frac{1}{\sigma_v^2}(x_n - y_n) + \frac{1}{\sigma_x^2}(x_n) = 0 \Leftrightarrow x_n = \frac{\sigma_x^2}{\sigma_v^2 + \sigma_x^2} y_n. \tag{12}$$

The estimation rule in (12) is also known as Wiener filter estimation and it was originally derived based on minimum mean-square error (MMSE) criterion. The MMSE criterion considers only the first- (i.e., mean) and second- (i.e., variance) order statistics. Since all higher order statistics are zero for Gaussian distributed signal and noise, both the MMSE and MAP criteria lead to the same estimation rule.

B. Gaussian Distributed Noise Only

Let us now consider that only noise is Gaussian distributed. For a Gaussian distributed noise, the term $E\{\ln p_v(\mathbf{y} - \mathbf{x})\}$ in (3) has no extremum. Therefore, the maximization in (3) is decided only by the term $E\{\ln p_x(\mathbf{x})\}$. According to our analysis in Section II, the maximum is achieved by using an ICA unmixing matrix of $\mathbf{x}$. The estimate of signal $\mathbf{x}$ can be obtained by using ICAMAP(gen) [i.e., (9)] by letting $W^v = W^x$. However, a simpler closed-form solution can be achieved as presented below. For clarity of presentation we denote the W^x by W. In this case, the MAP estimation rule from (8) can be expressed in the form of

$$\begin{aligned} \mathbf{x} &\leftarrow \arg\max_x \left(\sum_{n=1}^{N} \ln p(W_{n\cdot}(\mathbf{y} - \mathbf{x})) \right. \\ &\qquad \left. + \ln p(W_{n\cdot}\mathbf{x}) + 2\ln|\det(W)| \right) \\ &\Leftrightarrow \sum_{n=1}^{N} \arg\max_x (\ln p(W_{n\cdot}(\mathbf{y} - \mathbf{x})) + \ln p(W_{n\cdot}\mathbf{x})) \end{aligned} \tag{13}$$

where $W_{n\cdot} = W(n,:)$ denotes the nth row of matrix W.

We can perform the estimation in the independent space first and then transform the obtained estimate into the original space. Denote $W_n.\mathbf{y}$ as z_n and $W_n.\mathbf{x}$ as s_n. For each component, the estimate s_n can be obtained by solving

$$\begin{aligned} s_n &\leftarrow \arg\max_s (\ln p(z_n - s_n) + \ln p(s_n)) \\ &\Leftrightarrow \arg\min_s \left(\frac{1}{2\sigma^2}(z_n - s_n)^2 + f(s_n) \right). \end{aligned} \tag{14}$$

The minimization is equivalent to solving the following equation:

$$\frac{1}{\sigma^2}(s_n - z_n) + f'(s_n) = 0. \tag{15}$$

The MAP estimate can be obtained by $s = g(z)$. Although this may not have a closed form solution, the estimation function can be approximated as follows [6]:

$$s_n = \mathrm{sign}(z_n) \max(0, |z_n| - \sigma^2 |f'(z_n)|) \tag{16}$$

where $\mathrm{sign}(z_n)$ denotes the sign of z_n. The estimation rule in (16) is known as sparse code shrinkage estimation. This was originally proposed and employed for image enhancement in [6], and later used for speech enhancement in [7], [8] and extended to accommodate for a possible noise correlation in [10]. Unlike the above papers, we perform the modeling of signal distribution (i.e., function f) by using the generalized Gaussian model, as described in Section V. As such, we refer to the estimation rule in (16) as ICAMAP(ng-g) to reflect that the distribution of the signal is considered non-Gaussian but noise is Gaussian.

IV. Denoising Capability Analysis

In this section, we analyze the denoising capability of the estimation rule given by (3). In Bayesian estimation, the estimate of an unknown signal is achieved by minimizing a conditional risk R which is given by a cost function $C(\hat{x}, x)$ of estimating the true value of x as $\hat{x}$

$$\begin{aligned} R_x(\hat{x}) &= E\{C(\hat{x}, x)\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(\hat{x}, x) p(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} C(\hat{x}, x) p(x \mid y) dx \right] p(y) dy \end{aligned} \tag{17}$$

where y is the observed noisy data and x is the true value of data hidden in y.

As the *a priori* probability density of the observation y is often not known, we simply assign it to be a uniform density in our following derivation. Since the Bayesian risk is always positive, the minimization of $R_x(\hat{x})$ is achieved by selecting such $\hat{x}$ that for the given y the term in the bracket in (17) is minimum. Several forms of cost function can be chosen, which all depend on the problem to be solved. Suppose that in a given estimation problem we are not able to assign a particular cost function $C(\hat{x}, x)$, then a natural choice is a uniform cost function equal to zero between some certain small values $\pm\epsilon/2$ and uniform value outside that

$$C(\hat{x}, x) = \begin{cases} 0, & |x - \hat{x}| < \frac{\epsilon}{2} \\ \frac{1}{\epsilon}, & |x - \hat{x}|| > \frac{\epsilon}{2} \end{cases}. \tag{18}$$

Equation (17) then becomes

$$\begin{aligned} R_x(\hat{x}) &= E\left\{\int_{-\infty}^{\infty} C(\hat{x}, x) p(x \,|\, y) dx\right\} \\ &= E\left\{\frac{1}{\epsilon}\left[1 - \int_{\hat{x}-\frac{\epsilon}{2}}^{\hat{x}+\frac{\epsilon}{2}} p(x \,|\, y) dx\right]\right\} \\ &= E\left\{\frac{1}{\epsilon} - p(\hat{x} \,|\, y)\right\} \\ &= E\left\{\frac{1}{\epsilon} - \frac{p(y \,|\, \hat{x}) p(\hat{x})}{p(y)}\right\}. \end{aligned} \tag{19}$$

In the above derivation, the integral mean-value theorem was used. As $p(y)$ is considered to be uniformly distributed, it does not affect the risk minimization and can be omitted. Equation (19) shows that for any fixed ϵ, the larger the likelihood $E\{p(y \,|\, \hat{x}) p(\hat{x})\}$ is, the smaller the Bayesian risk will be. Without affecting the convergence, we consider the density functions to be compressed by a logarithm function. Without causing any confusion, we just use x to replace the estimate $\hat{x}$ in our following derivation.

As was discussed in Section II, the maximization of log likelihood can be solved within the ICA framework, and the maximum of the term $E\{\ln p_x(\mathbf{x})\}$ equals to negative value of the entropy $H(p_{x,A^*}(\mathbf{x}))$. Considering (5) and linear transformation of entropy given in [17], we obtain

$$H(p_{x,A^*}(\mathbf{x})) = H(p(\mathbf{s}_x)) + \ln|\det(A^*)|. \tag{20}$$

In the ICA framework, the value of $\ln|\det(A^*)|$ is equal to $-\ln|\det(U_x)|$, where U_x is the whitening matrix of $\mathbf{x}$. As such the maximum log likelihood can be expressed as

$$E\{\ln p_x(\mathbf{x})\}_{\max} = -H(p(\mathbf{s}_x)) + \ln|\det(U_x)|. \tag{21}$$

The entropy of $\mathbf{s}$ can be, by means of Edgeworth expansion [18], approximated as a sum of kurtosis and entropy of Gaussian variable $\mathbf{s}_{\text{gauss}}$ with the same covariance matrix as $\mathbf{s}_x$

$$H(p(\mathbf{s})) \approx H(p(\mathbf{s}_{\text{gauss}})) - \frac{1}{48} k(\mathbf{s}_x)^2 \tag{22}$$

where the entropy of Gaussian can be evaluated as [17]

$$H(p(\mathbf{s}_{\text{gauss}})) = \frac{N}{2}[1 + \log 2\pi] \tag{23}$$

where N is the dimension of $\mathbf{s}_{\text{gauss}}$.

Sum up (21) to (23), the maximum of the given variable's log likelihood can be approximated in the form of

$$\begin{aligned} E\{\ln p_x(\mathbf{x})\}_{\max} \approx &-\frac{N}{2}[1 + \ln 2\pi] \\ &+ \ln|\det(U_x)| + \frac{1}{48} k(\mathbf{s}_x)^2. \end{aligned} \tag{24}$$

For the term $E\{\ln p(\mathbf{y} \,|\, \mathbf{x})\}$, the probability $p(\mathbf{y} \,|\, \mathbf{x})$ is the probability of noise evaluated at point $\mathbf{y} - \mathbf{x}$, which we denote as $\mathbf{v}$. Similar expression as in (24) can be obtained for the maximum likelihood of this term

$$\begin{aligned} E\{\ln p_v(\mathbf{v})\}_{\max} \approx &-\frac{N}{2}[1 + \ln 2\pi] \\ &+ \ln|\det(U_v)| + \frac{1}{48} k(\mathbf{s}_v)^2. \end{aligned} \tag{25}$$

At this point, we have decomposed the value of log likelihood in (24) and (25) into the sum of three parts: the first term is a constant, the second term corresponds to second-order statistics, and the third term corresponds to higher-order statistics. To evaluate the denoising capability of the proposed algorithm for signals/noises of various statistical properties, variances of each given signal and each given noise are assumed to be fixed (i.e., SNR is fixed). The determinant terms in (24) and (25) can be considered constants, since the determinant of a whitening matrix corresponds to the determinant of its eigenvalue diagonal matrix, and eigenvalues are the variance of a given noise/signal. As such the maximum value of $E\{\ln p(\mathbf{y}|\mathbf{x}) p(\mathbf{x})\}$ can be approximated as

$$\begin{aligned} E\{\ln(p(\mathbf{y} \,|\, \mathbf{x}) p(\mathbf{x}))\}_{\max} \approx &\left(\frac{1}{48} k(\mathbf{s}_v)^2 + C_v\right) \\ &+ \left(\frac{1}{48} k(\mathbf{s}_x)^2 + C_x\right) \end{aligned} \tag{26}$$

where C_v and C_x denote constants corresponding to the sum of the first two terms in (24) and (25), respectively.

Considering (19) and (26), and the fact that the kurtosis is zero for a Gaussian distributed variable and its absolute value increases with increasing the deviation from the Gaussian distribution, the minimum of Bayes' risk denoted by $R_{x,v}$ will be in the order of

$$R_{(x_{\text{ng}}, v_{\text{ng}})} < R_{(x_{\text{ng}}, v_g)} < R_{(x_g, v_g)} \tag{27}$$

where the subscripts "ng" and "g" denote non-Gaussian and Gaussian distribution, respectively. Since a smaller value of risk means better denoising capability, the above inequality relationship expresses that the denoising capability improves with increasing the non-Gaussianity of signal/noise. Note that when both the signal and noise are Gaussian, it was shown in Section III-A that the MAP estimator leads to the best linear estimator (i.e., Wiener filter) and thus allowing nonlinearity in the estimation would not improve the performance.

The advantage of the proposed ICAMAP algorithms to other MAP algorithms based on using the wavelet [9], DCT and PCA

[11] can be illustrated by (27). Since the DCT and PCA transforms capture only second-order statistics of the signal, they cannot exploit the non-Gaussianity of signal/noise. The use of wavelet transform can exploit higher order statistics, however, since the basis functions are not constructed based on the actual signal the signal enhancement may not be optimal. As such the denoising capability of these algorithms will be between that of the Wiener filter and ICAMAP(ng–g) algorithms.

V. The Probability Density Model

Modeling of the density function of speech (and noise) is an important issue. The model should be a good fit to various degrees of non-Gaussianity and feasible.

A. Generalized Gaussian Model

We have adopted modeling based on the generalized Gaussian model (GGM), introduced in [14] and recently used in [19], as it can model well a wide-range of non-Gaussian distributions. The GGM is in a general form of

$$p(x \mid \mu, \delta, \beta) = \frac{\omega(\beta)}{\delta} \exp\left[-c(\beta)\left|\frac{x-\mu}{\delta}\right|^{2/(1+\beta)}\right] \tag{28}$$

where

$$c(\beta) = \left[\frac{\Gamma[3(1+\beta)/2]}{\Gamma[(1+\beta)/2]}\right]^{1/(1+\beta)} \tag{29}$$

$$\omega(\beta) = \frac{\Gamma[3(1+\beta)/2]^{1/2}}{(1+\beta)\Gamma[(1+\beta)/2]^{3/2}} \tag{30}$$

where Γ is gamma function.

The μ and δ denote the mean and standard deviation of the data, respectively. The parameter β controls the deviation of the distribution from Gaussian; by varying β, (28) can describe Gaussian, sub-Gaussian, and super-Gaussian distributions. For instance, when $\beta = 0$ the distribution is Gaussian and when $\beta = 1$ it is Laplacian; as $\beta \to -1$, the distribution becomes uniform and as $\beta \to \infty$, the distribution is a delta function.

For zero mean and unit variance variable, the differential of log probability density functions f' in (9) and (16) can be expressed as

$$f'(x) = \frac{2c(\beta)}{1+\beta}|x|^{2/(1+\beta)-1}. \tag{31}$$

B. Parameter Estimation

In our case, the signal is assumed to be zero mean and unit variance, the problem then reduces to the estimation of the value of β. Using the MAP method, the parameter β can be estimated based on samples of a given training signal $\mathbf{x} = \{x(t)\}_t$ as

$$\beta = \arg\max_{\beta} p(\beta \mid \mathbf{x}) \Leftrightarrow \arg\max_{\beta} p(\mathbf{x} \mid \beta) p(\beta) \tag{32}$$

where the data likelihood is

$$p(\mathbf{x} \mid \beta) = \prod_t \omega(\beta) \exp\left[-c(\beta)|x_t|^{2/(1+\beta)}\right] \tag{33}$$

and $p(\beta)$ defines the prior distribution of β.

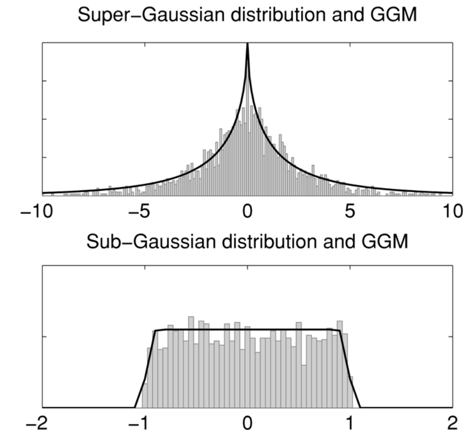

Fig. 1. Examples of data distributions and the corresponding estimated GGMs.

To model this prior distribution, the authors in [14] suggested to use a symmetric function with one adjustable parameter. This method however restricts the possible values for β to be within the range $(-1, 1)$. The authors in [19] proposed a more flexible model, which we have also adopted in our work. This models the beta prior using the Gamma distribution as

$$p(\beta) \sim \text{Gamma}(\beta \mid a, b) = \frac{b^a}{\Gamma(a)} \beta^{a-1} \exp(-b\beta) \tag{34}$$

where a and b are adjustable parameters. We used $a = b = 2$, which gives a broad prior distribution with a 95% density within the range of $[-0.5, 10.5]$.

The parameter β can be estimated based on (32) by applying a standard gradient ascent algorithm or some other more efficient algorithms. We have employed the DIRECT algorithm proposed in [20]. Our experiments showed that this algorithm estimates the β well and converges fast (normally in less than ten iterations). An example of histograms of super-Gaussian and sub-Gaussian data and the corresponding estimated GGMs are depicted in Fig. 1.

VI. Summary of the Proposed ICAMAP Algorithm

This section provides a summary of steps of the proposed ICA-based MAP enhancement algorithm in the general case (ICAMAP(gen)), as described in Section II.

1. Using two sets of data $\hat{\mathbf{v}}$ and $\hat{\mathbf{x}}$, which should have the same statistical properties as the noise $\mathbf{v}$ and signal $\mathbf{x}$, calculate the ICA transformation matrices W_v and W_x. This can be performed by using any of the existing ICA algorithms.
2. Estimate the generalized Gaussian density function from independent components $\mathbf{s}_v = W^v\hat{\mathbf{v}}$ and $\mathbf{s}_x = W^x\hat{\mathbf{x}}$ as described in Section V.
3. To perform the enhancement process of the observed noisy signal $\mathbf{y}$, set the step size λ and initial estimate of the clean signal $\hat{\mathbf{x}}$.
4. Apply the estimation rule (9) to update the estimate $\hat{\mathbf{x}}$.
5. Repeat the step 4 until the estimation process converges or maximum number of iterations is reached.

The steps of the proposed algorithm for a special case of Gaussian noise and non-Gaussian signal (i.e., ICAMAP(ng–g)) correspond to the scheme described in [6], with a difference that here we employed the GGM to model the signal distribution.

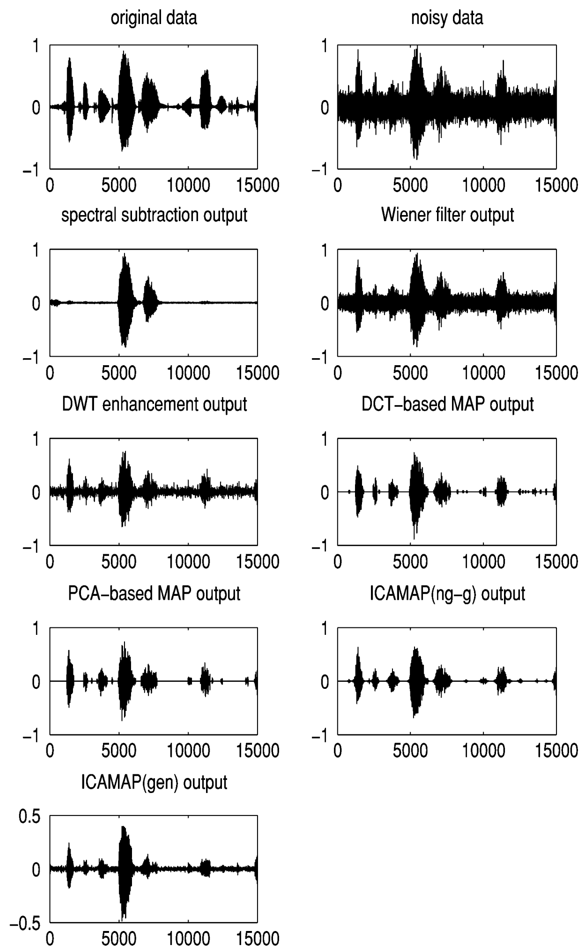

Fig. 2. Waveforms of female speech corrupted by Gaussian noise at SNR = 0 dB.

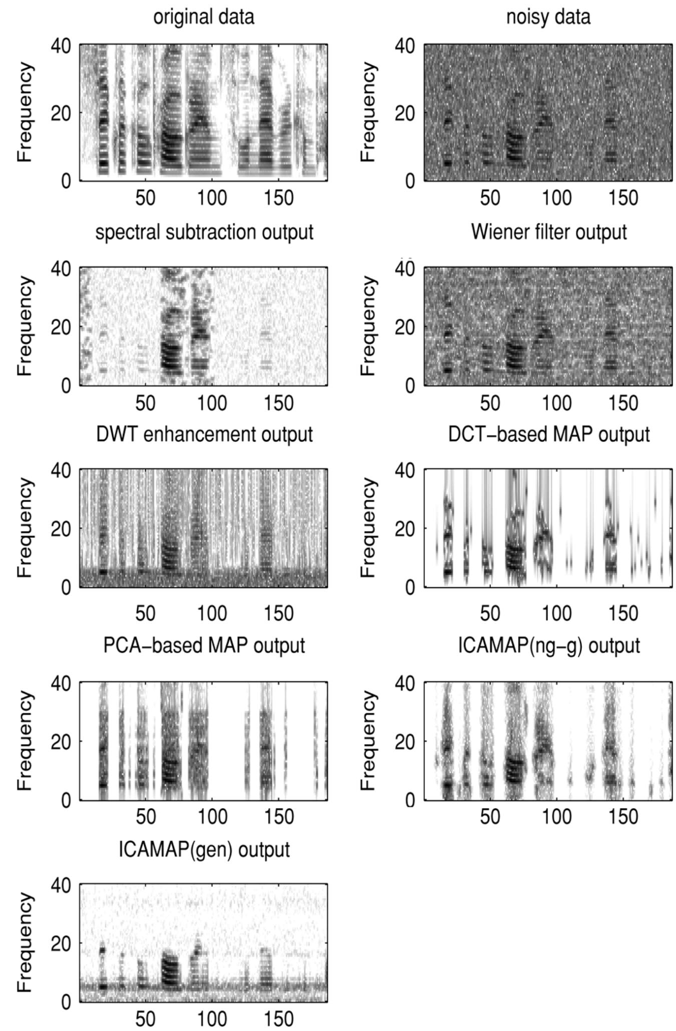

Fig. 3. Spectrograms of female speech corrupted by Gaussian noise at SNR = 0 dB.

VII. Experimental Evaluation

Experimental evaluation of the proposed algorithms is performed in terms of SNR and spectral distortion (SD) measure.

Let us denote the original clean speech and estimated speech signal by x and $\hat{x}$, respectively. The output SNR is then defined as

$$\mathrm{SNR}_{\mathrm{dB}} = 10\log_{10}\frac{\sum x^2}{\sum(x-\hat{x})^2}. \tag{35}$$

As the SNR itself may not provide enough information about the usefulness of an enhancement algorithm, we also provide evaluation by calculating spectral distortion (SD). The SD between signals x and $\hat{x}$ can be calculated as follows. First, in order to neglect the gain difference between the signals, each signal is normalized by performing $x = x/\|x\|$ and $\hat{x} = \hat{x}/\|\hat{x}\|$. Each signal is then divided into frames of 64 samples by using the rectangular window and each frame is appended by 192 zeros. The SD is then defined as an average difference between logarithm spectrum of each frame as

$$\mathrm{SD} = \frac{1}{4I}\sum_{i=1}^{I}\sum_{k=0}^{255} 20|\log_{10}|X(k,i)| - \log_{10}|\hat{X}(k,i)|| \tag{36}$$

where k and i are the frequency-index and frame-index, respectively.

The experimental evaluation was performed with speech signals selected from the TIMIT database, separately for each gender. The training set contained 30 sentences for each gender, which were randomly selected from the DR1 subset, and testing set contained five sentences of speaker fdaw0 (female) and mcpm0 (male); sentences from testing speakers were not used in the training set. Noisy speech signals were created by adding a noise to the clean speech signal at various SNRs. We used Gaussian noise and three non-Gaussian noises, Railway and Pub noise from the Noisex92 database, and Street noise from the Aurora2 database. Half of each noise signal was used to train the ICA basis functions and the corresponding GGM parameters of noise, while the other half was used to corrupt the test signals. All of the signals were sampled at 8 kHz. The signals were split into frames of 64 samples by using the rectangular window. The fast-ICA algorithm [17] was employed to estimate ICA basis functions based on the training data. The performance of the proposed algorithms ICAMAP(gen) and ICAMAP(ng–g) was compared with the spectral subtraction, Wiener filter, discrete wavelet transform (DWT) enhancement method proposed in [21] (Matlab function "wden" employing Symlets wavelet was used) and MAP algorithms based on (16) but using DCT, PCA transforms instead of ICA.

TABLE I
COMPARISON OF SNRs (IN dB) OF ENHANCED SIGNALS UNDER GAUSSIAN NOISE CORRUPTION

	Enhancement algorithm	Gaussian noise 0dB	Gaussian noise 5dB
Male	Spectral subtraction	2.3	6.2
	Wiener filter	1.5	7
	DWT enhancement	3.3	6.7
	DCT-based MAP	3	6.9
	PCA-based MAP	3.2	7
	ICAMAP(ng-g)	4	7.4
	ICAMAP(gen)	4.1	7.2
Female	Spectral subtraction	3.1	6.6
	Wiener filter	2.8	6.8
	DWT enhancement	3.6	7.1
	DCT-based MAP	2.9	6.7
	PCA-based MAP	3	7
	ICAMAP(ng-g)	3.8	7.6
	ICAMAP(gen)	3.3	7

TABLE II
COMPARISON OF SPECTRAL DISTORTION OF NOISY SIGNALS (INPUT SD) AND ENHANCED SIGNALS UNDER GAUSSIAN NOISE CORRUPTION

	Enhancement algorithm	Gaussian noise 0dB	Gaussian noise 5dB
Male	Input SD	12.9	10.7
	Spectral subtraction	10.5	9.4
	Wiener filter	11.7	9
	DWT enhancement	9.5	8.4
	DCT-based MAP	12	10.5
	PCA-based MAP	11.2	9.3
	ICAMAP(ng-g)	9.6	8.2
	ICAMAP(gen)	9.3	8
Female	Input SD	14.4	12.1
	Spectral subtraction	12	9.5
	Wiener filter	13.1	10
	DWT enhancement	11.1	9.3
	DCT-based MAP	11.8	10.3
	PCA-based MAP	11.5	9.8
	ICAMAP(ng-g)	11	9.2
	ICAMAP(gen)	10.8	8.9

First the evaluation is performed for a Gaussian noise corruption. An example of clean, noisy and enhanced speech signals and corresponding spectrograms are depicted in Figs. 2 and 3, respectively. It can be observed in the time-domain representation that the enhanced speech signals obtained by all the algorithms but the Wiener filter and DWT have a low residual noise level. The output of spectral subtraction is overdenoised, i.e., some of speech signal have been removed. From the spectrogram representation it can be seen that the result of ICAMAP(ng–g) is close to the original clean signal. The ICAMAP(gen) preserves the harmonic structure slightly better than the ICAMAP(ng–g), however, it suppresses high frequency components. It can also be seen that the spectrograms of output of the DCT and PCA-based MAP algorithms are similar to the ICAMAP(ng–g). However, the evaluation in terms of SNR and SD as presented below shows better performance of the ICAMAP(ng–g) than the other two MAP algorithms. The obtained results in terms of SNR and SD are presented in the Tables I and II, respectively. It can be seen that both of the ICAMAP algorithms in most cases outperformed the other presented algorithms. The performance of the ICAMAP(gen) and ICAMAP(ng–g) is similar, which is an expected outcome since the ICAMAP(ng–g) is a special case of ICAMAP(gen) under the assumption of Gaussian distributed noise. The slight difference between the ICAMAP algorithms may be due to different optimization methods they used.

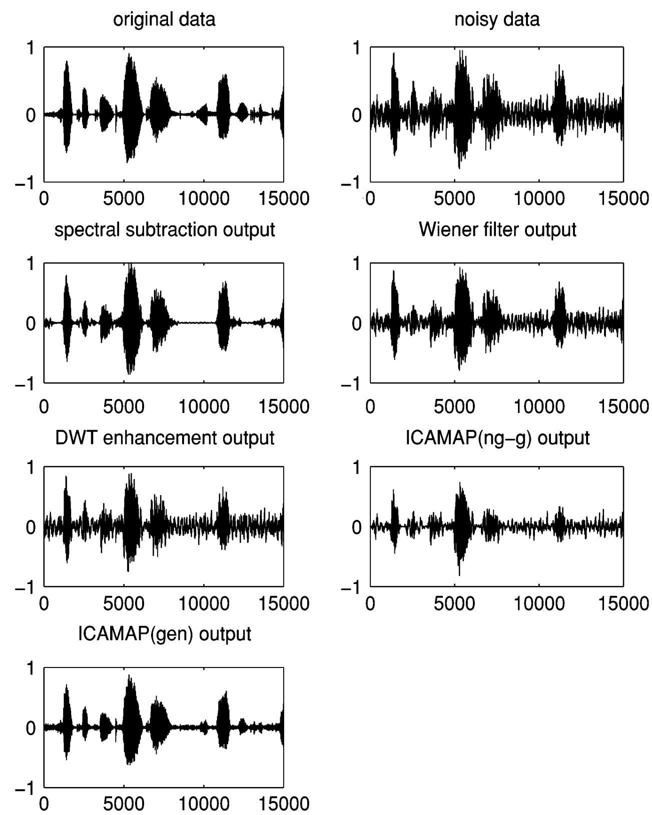

Fig. 4. Waveforms of female speech corrupted by Railway noise at SNR = 0 dB.

Here experimental results obtained for non-Gaussian noise corruption, i.e., Pub, Railway station, and Street noises, are presented. Figs. 4 and 5 depict example of signals and corresponding spectrograms, respectively. It can be observed in the time-domain representation that the ICAMAP(gen) enhanced speech signal is the closest to the original signal compared to the other four algorithms. The spectrogram representation shows that while the ICAMAP(gen), spectral subtraction, Wiener filter and DWT enhancement algorithms preserve the harmonic structure better than ICAMAP(ng–g), the ICAMAP(gen) achieves significantly better result than the other four algorithms, which is indeed very similar to the original clean signal. The results for non-Gaussian noises in terms of SNR and SD are presented in Tables III and IV, respectively. It can be seen that the proposed ICAMAP(gen) algorithm achieved significant improvement over the spectral subtraction, Wiener filter, DWT and ICAMAP(ng-g) algorithms for both male and female speakers at both SNRs for each noise.

Note that the computational requirements of the ICAMAP(ng–g) is effectively the same as the other two MAP algorithms (DCT, PCA) as the ICA matrix is obtained during the training mode. Due to the gradient-based iterative process the proposed ICAMAP(gen) algorithm has considerably higher computational requirements than the closed-form solutions of the ICAMAP(ng–g) algorithm, and it may be sensitive to the parameters used within the estimation process. Nevertheless, the results obtained demonstrate that the ICAMAP(gen) algorithm, i.e., employment of higher-order statistics in the

TABLE III
COMPARISON OF SNRS (IN dB) OF ENHANCED SIGNALS UNDER NON-GAUSSIAN NOISE CORRUPTION

	Enhancement algorithm	Pub		Railway		Street	
		0dB	5dB	0dB	5dB	0dB	5dB
Male	Spectral subtraction	1.8	6	5.6	8.9	4.9	7
	Wiener filter	0.8	5.6	2.6	6.5	1.8	5.8
	DWT enhancement	0.5	5	0.9	5.1	0.5	5
	ICAMAP(ng-g)	1.3	5.1	1.1	5.3	1.5	5.1
	ICAMAP(gen)	6.9	10.2	7	10.9	6.2	9
Female	Spectral subtraction	3.2	7.4	6.8	8.6	4.5	7.8
	Wiener filter	1.9	7	3.7	8.3	2	6.9
	DWT enhancement	0.4	5	0.8	4.9	0.3	5.1
	ICAMAP(ng-g)	2.8	5.8	2.5	5.7	2.5	5.4
	ICAMAP(gen)	5.9	9.6	10.8	12.7	6.4	9.6

TABLE IV
COMPARISON OF SPECTRAL DISTORTION OF NOISY SIGNALS (INPUT SD) AND ENHANCED SIGNALS UNDER NON-GAUSSIAN NOISE CORRUPTION

	Enhancement algorithm	Pub		Railway		Street	
		0dB	5dB	0dB	5dB	0dB	5dB
Male	Input SD	7.6	5.8	7.9	6.2	8	6.3
	Spectral subtraction	9	8.4	7.6	6.5	8.9	7.3
	Wiener filter	8.2	6.7	7.7	6.4	8.4	6.2
	DWT enhancement	8.5	6.9	8.2	6.7	8.4	6.9
	ICAMAP(ng-g)	12	9.7	10.1	8.6	10.4	8.9
	ICAMAP(gen)	6.6	5.5	7.2	5.4	7.4	5.6
Female	Input SD	8.8	6.9	9.1	7.2	9.8	7.2
	Spectral subtraction	8.9	7.1	8.7	6.9	8.6	7
	Wiener filter	9.1	7.5	8.6	7.1	8.4	6.9
	DWT enhancement	9.1	7.6	9.1	7.4	9.3	7.5
	ICAMAP(ng-g)	12	9.6	9.5	8.9	10.2	9.1
	ICAMAP(gen)	6.8	6	6.4	5.3	7.2	6.4

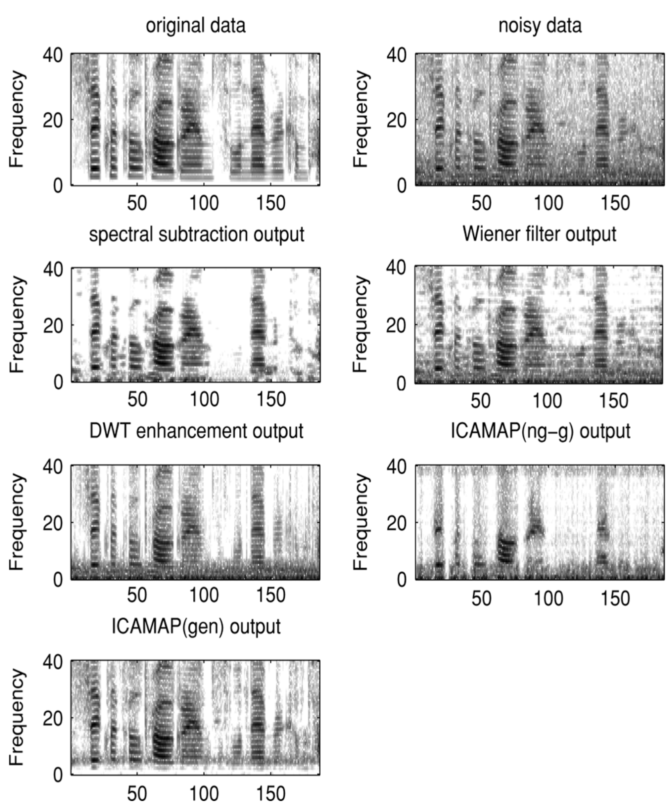

Fig. 5. Spectrograms of female speech corrupted by Railway noise at SNR = 0 dB.

estimation rule, can provide significantly better enhancement performance.

VIII. CONCLUSION

In this paper, we presented a novel framework for speech signal enhancement based on the MAP algorithm employed in the ICA space. It was demonstrated that, within the linear transformation domain, the best estimate can be obtained by employing individual ICA transformations for signal and noise. It was also shown that the proposed framework under the assumption of both the signal and noise being Gaussian distributed leads to the Wiener filter and under the assumption of only noise being Gaussian leads to the sparse code shrinkage. Our theoretical analysis of the denoising capability of the proposed algorithm demonstrated that it is most effective for non-Gaussian signals corrupted by a non-Gaussian noise while the case of a Gaussian signal and Gaussian noise is the most difficult to deal with. The modeling of the signal and noise was performed by using the generalized Gaussian model that provides a flexibility to account for a wide-range of non-Gaussian distributions. Experiments were performed on speech data corrupted by Gaussian and non-Gaussian noises and the evaluation was performed in terms of SNR and spectral distortion measure. The experimental results, confirming our theoretical analyses, showed significant performance improvements over conventional speech enhancement techniques.

REFERENCES

[1] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, no. 2, pp. 113–120, Apr. 1979.

[2] C. He and G. Zweig, "Adaptive two-band spectral subtraction with multi-window spectral estimation," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Phoenix, AZ, 1999, vol. 2, pp. 793–796.

[3] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Orlando, FL, May 2002, vol. 4, pp. 4164–4167.

[4] J. Lim and A. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, pp. 1586–1604, 1979.

[5] I. Y. Soon and S. N. Koh, "Low distortion speech enhancement," *Proc. Inst. Elec. Eng.*, vol. 147, no. 3, pp. 247–253, 2000.

[6] A. Hyvarinen, "Sparse code shrinkage: Denoising of nonGaussian data by maximum likelihood estimation," *Neural Computat.*, vol. 11, no. 7, pp. 1739–1768, 1999.

[7] J.-H. Lee, H.-Y. Jung, T.-W. Lee, and S.-Y. Lee, "Speech enhancement with MAP estimation and ICA-based speech features," *Electron. Lett.*, vol. 36, pp. 1506–1507, 2000.

[8] I. Potamitis, N. Fakotakis, G. Kokkinakis, and , "Speech enhancement using the sparse code shrinkage technique," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Salt Lake City, UT, 2001, pp. 621–624.

[9] H. Xie, L. E. Pierce, and F. T. Ulaby, "SAR speckle reduction using wavelet denoising and Markov random field modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2196–2212, 2002.

[10] L. Hong, J. Rosca, and R. Balan, "Bayesian single channel speech enhancement exploiting sparseness in the ICA domain," presented at the Eur. Signal Process. Conf. (EUSIPCO), Vienna, Austria, Sep. 6–10, 2004.

[11] S. Gazor and W. Zhang, "Speech enhancement employing Laplacian-Gaussian mixture," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 896–904, Sep. 2005.

[12] T. Letter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model," *EURASIP J. Appl. Signal Process.*, pp. 1110–1126, 2005.

[13] C. Breithaupt and R. Martin, "MMSE estimation of magnitude-squared DFT coefficients with super-Gaussian priors," in *Proc. Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, Hong Kong, China, Apr. 2003, vol. 1, pp. 896–899.

[14] G. Box and G. Tiao, *Bayesian Inference in Statistical Analysis*. New York: Wiley, 1973.

[15] J.-F. Cardoso, "Infomax and maximum likelihood for blind source separation," *IEEE Signal Process. Lett.*, vol. 4, no. 4, pp. 112–114, Apr. 1997.

[16] X. Zou, P. Jancovic, and J. Liu, "The effectiveness of ICA-based representation: Application to speech feature extraction for noise robust speaker recognition," presented at the Eur. Signal Process. Conf. (EUSIPCO), Florence, Italy, Sep. 4–8, 2006.

[17] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

[18] J.-F. Cardoso, "High-order constrasts for independent component analysis," *Neural Computat.*, vol. 11, no. 1, pp. 157–192, 1999.

[19] T. W. Lee and G. J. Lewicki, "The generalized Gaussian mixture model using ICA," in *Proc. Int. Workshop Independent Compon. Anal.*, Jun. 2000, pp. 239–244.

[20] B. Mattias and H. Kenneth, "Global optimization using the DIRECT algorithm in Matlab," *Adv. Model. Optimiz.*, vol. 1, no. 2, 1999.

[21] D. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 613–627, 1995.

Xin Zou received the B.S. degree in electronics information science and engineering from Shandong University, China, in 2002. He is currently working towards the Ph.D. degree at the Department of Electronic, Electrical and Computer Engineering, University of Birmingham, Birmingham, U.K.

His research interests include digital signal processing, oriented towards speech and image enhancement, feature extraction, independent component analysis, and blind source separation.

Peter Jančovič (M'06) received the B.Sc. and M.Sc. degrees in information technology from the Slovak University of Technology, Bratislava, Slovakia, in 1997 and 1999, respectively, and the Ph.D. degree in the field of noise robust speech recognition from the School of Computer Science, Queen's University Belfast, Northern Ireland, U.K., in 2002.

He is currently a Lecturer with the Department of Electronic, Electrical and Computer Engineering, University of Birmingham, U.K. He was on a study visit with the School of Computer Science, Queens University Belfast, from July to September 1999 and with the Department of Signal Theory and Communication, Polytechnical University of Catalonia, Barcelona, Spain, from April to July 1997. His current research interests include noise-robust automatic speech and speaker recognition, speech signal enhancement, independent component analysis, time-frequency representation, data analysis, and pattern recognition.

Dr. Jančovič has served as an international program committee member of the World Congress on Engineering (2007) and IASTED conference Signal Processing, Pattern Recognition, and Applications in 2007 and 2008. He is a member of the IET.

Ju Liu (M'01) received the B.S. and M.S. degrees, both in electronic engineering, from Shandong University (SDU), Jinan, China, in 1986 and 1989, respectively, and the Ph.D. degree in signal processing from Southeast University (SEU), Nanjing, China, in 2000.

Since July 1989, he has been with the Department of Electronic Engineering of the SDU. He had been a Teaching Assistant (1989–1992), Lecturer (1993–1998), Associate Professor (1998–2000), and Full Professor since 2000 with Shandong University. From July 2002 to December 2003, he was a Visiting Professor with the Department of Signal Theory and Communication, Polytechnic University of Catalonia (UPC) and Telecommunications Technological Centre of Catalonia (CTTC), Barcelona, Spain. From November 2005 to January 2006, he was a DAAD researcher with the Department of Communication Engineering, University of Bremen, and the Department of Communication Systems, University of Duisburg-Essen, Germany. He has also been a Visiting Professor for a short time with Korea and Japan in June and August 2006, respectively. He is currently Chair Professor of the Department of Communication Engineering and Vice Director of the academic committee of the School of Information Science and Engineering, Shandong University, and Director of Texas Instruments (TI)-Shandong University DSPs Laboratory. He is the author of more than 100 journal and conference papers. His present research interests include space-time processing in wireless communication, blind signal separation and independent component analysis, multimedia communication, and DSP applications.

Dr. Liu is also on the editorial committee of the *Journal of Circuits and Systems* and the *Journal of Data Acquisition and Processing* in China. He is the holder of the Program for New Century Excellent Talents in University (NCET) of China, and the holder of three national and local academic awards in blind signal processing.

Münevver Köküer (S'93–M'95) received the B.Sc. degree in electrical engineering from Gazi University, Ankara, Turkey, and the M.Sc. degree in electrical-electronic engineering from Anadolu University, Eskisehir, Turkey, and the Ph.D. degree in the field of image processing from the Department of Electronic Systems Engineering, University of Essex, U.K., in 1994,

From 1994 to 2001, she was with Anadolu University, Turkey, where she was an Associate Professor. From 2000 to 2003, she had been a Research Fellow at Queen's University Belfast, U.K. She was Invited Lecturer with the Summer Session Programme of International Space University (ISU), Bremen, Germany, in 2001 and Los Angeles, CA, in 2002. From 2003 to 2006, she was with the Biomedical Computing Department, Faculty of Engineering and Computing, Coventry University, U.K. She is currently a Research Fellow with the Department of Electronic, Electrical and Computer Engineering, University of Birmingham, U.K. Her current research interests include speech pattern processing, data analysis, and image processing.

Dr. Köküer served as an Assistant Editor for the journal *Neurocomputing* (Elsevier) from 2001 to 2002. She is a member of TMMOB.