Frequency-Domain Pearson Distribution Approach for Independent Component Analysis (FD-Pearson-ICA) in Blind Source Separation

Hiroko Kato Solvang, *Member, IEEE*, Yuichi Nagahara, Shoko Araki, *Member, IEEE*, Hiroshi Sawada, *Senior Member, IEEE*, and Shoji Makino, *Fellow, IEEE*

***Abstract*—In frequency-domain blind source separation (BSS) for speech with independent component analysis (ICA), a practical parametric Pearson distribution system is used to model the distribution of frequency-domain source signals. ICA adaptation rules have a score function determined by an approximated signal distribution. Approximation based on the data may produce better separation performance than we can obtain with ICA. Previously, conventional hyperbolic tangent ($tanh$) or generalized Gaussian distribution (GGD) was uniformly applied to the score function for all frequency bins, even though a wideband speech signal has different distributions at different frequencies. To deal with this, we propose modeling the signal distribution at each frequency by adopting a parametric Pearson distribution and employing it to optimize the separation matrix in the ICA learning process. The score function is estimated by the appropriate Pearson distribution parameters for each frequency bin. We devised three methods for Pearson distribution parameter estimation and conducted separation experiments with real speech signals convolved with actual room impulse responses ($T_{60} = 130$ ms). Our experimental results show that the proposed frequency-domain Pearson-ICA (FD-Pearson-ICA) adapted well to the characteristics of frequency-domain source signals. By applying the FD-Pearson-ICA performance, the signal-to-interference ratio significantly improved by around 2–3 dB compared with conventional nonlinear functions. Even if the signal-to-interference ratio (SIR) values of FD-Pearson-ICA were poor, the performance based on a disparity measure between the true score function and estimated parametric score function clearly showed the advantage of FD-Pearson-ICA. Furthermore, we confirmed the optimum of the proposed approach for/optimized the proposed approach as regards separation performance. By combining individual distribution parameters directly estimated at low frequency with the appropriate parameters optimized at high frequency, it was possible to both reasonably improve the FD-Pearson-ICA performance without any significant increase in the computational burden by comparison with conventional nonlinear functions.**

***Index Terms*—Convolutive mixtures, Kurtosis, Pearson types I, IV, and VI, score function, skewness, speech separation.**

Manuscript received September 29, 2006; revised October 17, 2008. Current version published March 27, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Rudolf Rabenstein.

H. K. Solvang is with Norwegian Radium Hospital, Rikshospitalet University Hospital, Montebello 0310 Oslo, Norway, and also with the Department of Biostatistics, Institute of Basic Medical Science, University of Oslo, NO-0316 Oslo Norway (e-mail: hiroko.solvang@rr-research.no).

Y. Nagahara is with Meiji University, Tokyo 101-8301, Japan (e-mail: nagahara@kisc.meiji.ac.jp).

S. Araki, H. Sawada, and S. Makino are with NTT Communication Science Laboratories, Kyoto 619-0237, Japan (e-mail: shoko@cslab.kecl.ntt.co.jp; sawada@cslab.kecl.ntt.co.jp; maki@cslab.kecl.ntt.co.jp).

Digital Object Identifier 10.1109/TASL.2008.2011527

I. Introduction

Blind source separation (BSS) estimates original source signals by using only the information provided by observed mixtures. Independent component analysis (ICA) [1]–[3] is one of the main statistical methods of BSS. The BSS of speech signals, which is the main topic of this contribution, has a wide range of applications, including robust noise/speech recognition, hands-free telecommunication systems, and more comfortable hearing aids.

This paper considers the BSS of speech signals in real environments, namely the BSS of convolutive mixtures. In a real environment, speech signals are recorded along with their reverberation. To separate such complicated mixtures, signals are usually converted into the frequency domain to form instantaneous mixture problems in each frequency bin [4]–[10], and this is called frequency-domain BSS. Frequency-domain BSS employs complex-valued ICA for instantaneous mixtures at each frequency.

An ICA learning rule generally includes the estimation of the score function [1]–[3]. For instance, $\tanh$ is an activation function used as an estimate of a score function. In fact, there is a connection between the activation function and the source prior in terms of maximum-likelihood (ML) estimation terms. [11] demonstrated the connection between ML-ICA, Natural Gradient and the FastICA algorithm [12], [13] and showed that the actual score function in FastICA can also be interpreted as a function that incorporates source prior information. As pointed out in [11], the selection of the score function performs quite important role in the ICA algorithm, and the score function is deeply related to the source priors. To obtain better separation performance, we must find appropriate source distributions for each frequency to realize a more suitable score function. Since the distributions are unknown in a blind scenario, approximated distributions are utilized. For speech separation, a super-Gaussian distribution has been uniformly used as the score function in all frequency bins, as seen in Fast ICA, which is one of the most widely used algorithms. To obtain a more efficiently converging version of FastICA, [14] used the constraint for the residual error variance. Then [15] adapted the shape of the source distribution to the data. When looking at the distributions of a speech signal at different frequencies, they are in fact *not similar* because they are fat-tailed and skewed according to each sequence. Therefore, it is preferable to model an appropriate distribution for each frequency bin.

1558-7916/$25.00 © 2009 IEEE

These various fat-tailed and skewed speech distribution shapes resemble the distribution shapes of the Pearson distribution system [16], which includes several distribution types for modeling various source distributions. In fact, a Pearson distribution system applied to ICA (Pearson-ICA) has been studied [17], and this approach achieved better separation performance than such conventional nonlinear functions as $\tanh$. Furthermore, a nonparametric ICA approach to estimating the source distribution was proposed, and its separation performance was compared with those of several methods including Pearson-ICA, Fast-ICA, and Kernel-ICA [18]. However, [17] and [18] were performed in the time-domain and used artificial data. In [17], Pearson-ICA was employed to solve the instantaneous BSS problem of artificial data, but Pearson-ICA for convolutive mixtures of speech data (i.e., where delay and filtering are considered) has never been studied. On the other hand, a generalized Gaussian distribution (GGD)-based nonlinear score function was employed for time-domain [19] and frequency-domain [20] speech signals. Since the shape parameters of GGD can adapt to the distribution shape, the approach seemed more flexible than a uniform application of a super-Gaussian distribution; however, these approaches were always applied to the time domain [19] or applied uniformly to all frequency bins [20]. Another problem with the GGD approach is that it cannot model a skewed distribution, which sometimes appears for speech signals in the frequency domain.

Leaving aside such problems, we focus on a more central issue: the convolutive BSS problem. As a solution to this practical problem, we study Pearson-ICA for the frequency-domain BSS of speech signals, which can deal with a more practical issue, namely the convolutive BSS problem. Such a frequency-domain BSS technique, using a Pearson distribution that adapts to the actual data distribution shape, has yet to be developed. Therefore, this article proposes our approach for applying the Pearson distribution system to frequency-domain BSS (FD-Pearson-ICA, Frequency-domain Pearson-ICA). We adapt appropriate Pearson distribution types to the individual distribution shape of each frequency.

This paper is organized as follows: Section II introduces the basic framework of the speech BSS that we handle. Section III outlines a practical parametric Pearson distribution system that involves applications with real speech data. Section IV introduces our proposed blind source separation methods. Section V describes experimental methods and the results of actual data analysis, based on a performance evaluation in terms of the signal-to-interference ratio (SIR). In the cases where the performance comparison obtained by using SIR was unclear, a disparity measure was applied to compare the parametric score functions for conventional $\tanh$, *GGD* and the proposed FD-Pearson-ICA with a true score function. In addition, we discuss the computational time problem for running programs, improving separation performance, and developing more efficiently expanded methods. Our conclusions are provided in Section VI.

II. BSS of Speech

A. Problem Description

We consider the BSS of speech signals observed in actual environments, i.e., the BSS of convolutive mixtures of speech. In such environments, N source signals $s_i(n)$ are observed with their reverberant components at M sensors. Therefore, observations are modeled as convolutive mixtures

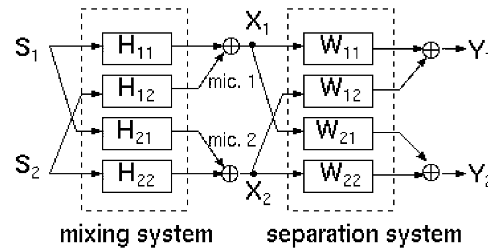

Fig. 1. Frequency-domain speech BSS system ($N = M = 2$).

$$x_j(n) = \sum_{i=1}^{N}\sum_{p=1}^{P} h_{ji}(p)s_i(n-p+1) \quad j = 1,\cdots,M \tag{1}$$

where $h_{ji}(n)$ is the P-taps impulse response from source i to sensor j. Our goal is to obtain separated signals $y_k(n)$ $k = 1,\cdots,N$ using only the information provided by observations $x_j(n)$. In this paper, we deal with the case where $N = M = 2$ (Fig. 1). An investigation of the performance with different numbers of sources and sensors is beyond the scope of this paper, although it would be easy to expand our proposed method for $N \geq M$.

This paper employs a frequency-domain approach for converting our problem into a linear instantaneous mixture at each frequency. In the frequency domain, mixtures (1) are modeled as

$$X_j(f,m) = \sum_{i=1}^{N} H_{ji}(f)S_i(f,m) \tag{2}$$

where f denotes a frequency and m is the frame index. With matrices, (2) can be written as

$$\mathbf{X}(f,m) = \mathbf{H}(f)\mathbf{S}(f,m)$$

where $\mathbf{H}(f)$ is an $M \times N$ mixing matrix whose ji component is a transfer function from source i to sensor j, and $\mathbf{S}(f,m) = [S_1(f,m),\cdots,S_N(f,m)]^T$ and $\mathbf{X}(f,m) = [X_1(f,m),\cdots,X_M(f,m)]^T$ denote the short-time Fourier transform (STFT) of sources and observed signals, respectively. In a blind scenario, $\mathbf{H}(f)$ and $\mathbf{S}(f,m)$ are unknown.

B. Previous Method

The separation process can be formulated at each frequency f

$$\mathbf{Y}(f,m) = \mathbf{W}(f)\mathbf{X}(f,m) \tag{3}$$

where $\mathbf{Y}(f,m) = [Y_1(f,m),\cdots,Y_N(f,m)]^T$ is the estimated source signal vector and $\mathbf{W}(f)$ is a separation matrix. $\mathbf{W}(f)$ is determined so that $Y_1(f,m),\cdots,Y_N(f,m)$ become mutually independent using ICA. After obtaining separated signals (3) and properly aligning the permutation and scaling ambiguities, we convert the frequency-domain signal $Y_k(f,m)$ into a time-domain signal by using inverse STFT.

The separation matrix is independently estimated at each frequency. An algorithm based on the natural gradient [21] is widely used. The adaptation rule of the ith iteration is

$$\mathbf{W}_{i+1}(f)=\mathbf{W}_i(f)+\eta\left[\mathbf{I}-\left\langle\Phi\left(\mathbf{Y}(f,m)\right)\mathbf{Y}^H(f,m)\right\rangle\right]\cdot\mathbf{W}_i(f) \quad (4)$$

where $\langle\mathbf{Y}(f,m)\rangle$ denotes an average with respect to m, H represents the transpose conjugate, and η is the adaptation step size. Here, $\Phi(\cdot)$ indicates the score function. If the source distributions p are known, score functions are defined as [1]–[4], [8]

$$\Phi(Y) = -\frac{p'\left(|Y|\right)}{p\left(|Y|\right)}\exp(j\angle Y) \quad (5)$$

where $p'(x) = dp(x)/dx$, Y is a complex number, $|\cdot|$ indicates the absolute value, and $\angle Y$ is the argument. In blind separation, however, the source distribution cannot be obtained *a priori*, and the score function is approximated by a nonlinear function. The score function $\tanh$ is widely used for speech separation because speech signals have a super-Gaussian distribution [1]–[3]

$$\Phi(Y) = \tanh\left(g|Y|\right)\exp(j\angle Y) \quad (6)$$

where g indicates a shape parameter.

With conventional GGD [20], the score function is represented by

$$\Phi(Y) = |Y|^{\beta-1}\exp(j\angle Y) \quad (7)$$

where β indicates the shape parameter. With a GGD, a Laplacian distribution whose speech closely follows it is defined as $\beta = 1$, a standard Gaussian distribution as $\beta = 2$, and a Gamma distribution as $\beta = 0.5$. Previous methods uniformly applied $\tanh$ and GGD [20] to all frequencies f. However, frequency-domain speech signals have various distributions at different frequencies. As references to express differences in distributions for different frequencies, the upper panels in Fig. 2(a) and the three panels in Fig. 2(b) show data histograms of absolute values of their STFT at frequency bins for f = 30, 2, 6, in Fig. 2(a), 150, 300, and 400, in Fig. 2(b), respectively. Here, the STFT frame size is 512, and the sampling rate is 8 kHz. Each figure in the upper panels of Fig. 2(a) describes a different distribution. With Fig. 2(b), even though it appears to show similar J-shaped figures, the heights and tails of the distributions are slightly different. Moreover, the distribution can also depend on the speakers. Therefore, it is inappropriate to apply a single score function to all frequencies/speakers in real source separation stages. To obtain good separation performance, we approximate appropriate source distributions frequency by frequency to model a more suitable score function. In this paper, we propose modeling the signal distribution and the score function at each frequency by a Pearson distribution, which is introduced in the next section.

III. Practical Approach With Pearson Distribution System

To obtain a more suitable score function, we applied the Pearson distribution system, which is widely used to model various source distributions. Pearson [16] defined the following differential equation related to probability density function $p(x)$:

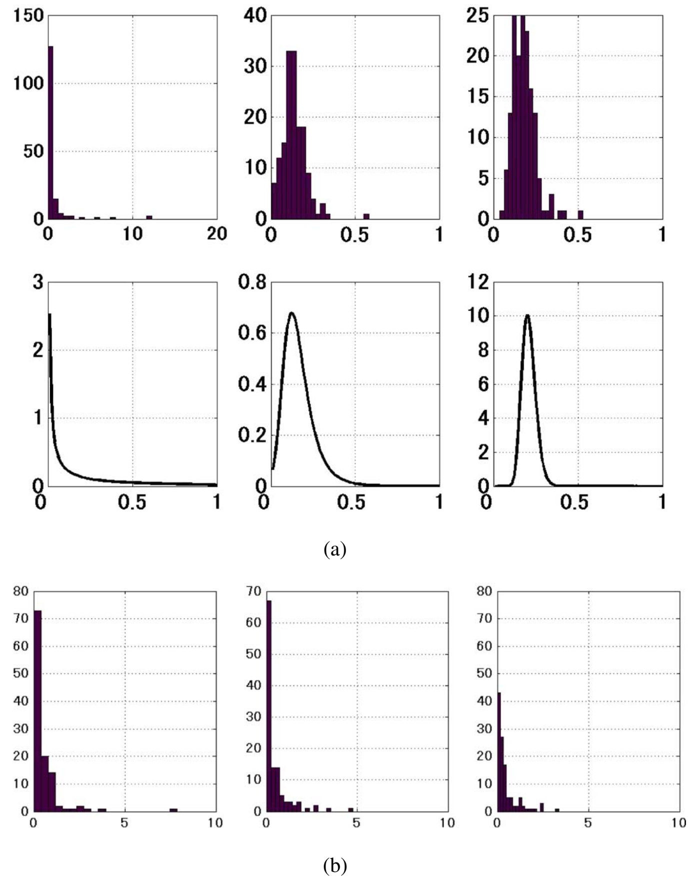

Fig. 2 (a) Upper three panels: Histograms of the STFT frame (frame size is 512 and sampling rate is 8 kHz). Estimated κ values specified the left, middle, and right histograms as Pearson Types I, IV, and VI distributions, respectively. Horizontal axis: $|Y(f,y)|$, vertical axis: frequencies. Lower three panels: the pdf curves using estimated parameters for Types I (left), IV (middle), and VI (right), respectively. (b) Histograms for frequency bins $f = 150, 300$, and 400. The STFT frame size is 512, and the sampling rate is 8 kHz. The patterns are Type I.

$$-\frac{p'(x)}{p(x)} = \frac{b_0 + b_1 x}{c_0 + c_1 x + c_2 x^2}. \quad (8)$$

Since we have to handle complex random variables, we modify (8) as

$$-\frac{p'\left(|Y|\right)}{p\left(|Y|\right)} = \frac{b_0 + b_1|Y|}{c_0 + c_1|Y| + c_2|Y|^2}. \quad (9)$$

Note that form (9) corresponds to the score function (5) of ICA and we obtain the following score function:

$$\Phi(Y) = \frac{b_0 + b_1|Y|}{c_0 + c_1|Y| + c_2|Y|^2}\exp(j\angle Y). \quad (10)$$

That is, if the coefficients of (9) can be estimated by an appropriate method through the observed data in each frequency, we can obtain a score function to approximate the source distribution at each frequency.

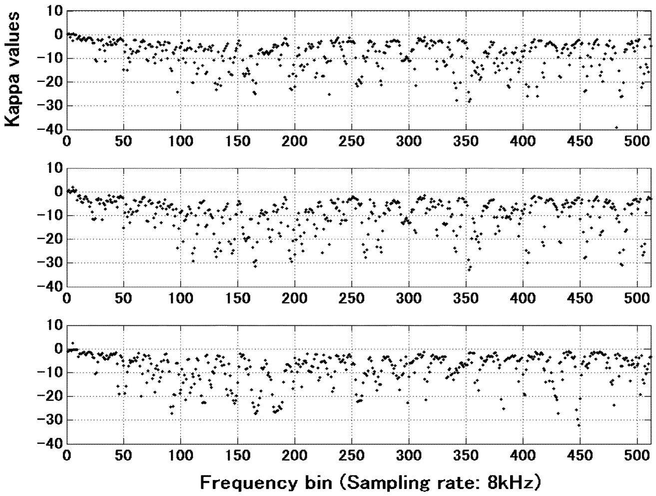

Fig. 3. κ values calculated by moments of STFT speech data. The top, second, and third panels indicate calculated κ values for combined female and male speech data, female speech data, and male speech data. The STFT frame size is 512, and the sampling rate is 8 kHz.

The Pearson distribution system mainly employs seven distribution types, although there are actually 12 types. A practical approach that uses all types of distribution is reported in [22]–[24], and its implementation to general data thus became simpler. First, to discriminate the Pearson type for the given data, [23] introduced a useful parameter κ

$$\kappa=\frac{(\mathrm{Skew})^2(\mathrm{Kurt}+3)^2}{4\left(2\times\mathrm{Kurt}-3\times(\mathrm{Skew})^2-6\right)\left(4\times\mathrm{Kurt}-3\times(\mathrm{Skew})^2\right)} \tag{11}$$

where $\mathrm{Skew}=\mathrm{E}[(x-\mathrm{E}\{x\})^3]/\mathrm{E}[(x-\mathrm{E}[x])^2]^{3/2}$ and $\mathrm{Kurt}=\mathrm{E}[(x-\mathrm{E}[x])^4]/\mathrm{E}[(x-\mathrm{E}[x])^2]^2$ for random variable x ($\mathrm{E}\{\cdot\}$ indicates the expectation of $\cdot$).

According to [23], the types for $\kappa<0$, $0<\kappa<1$, and $1<\kappa$ are I, IV, and VI, respectively. In Fig. 2(a), the upper and lower panels on the left show the data histogram and the Type I probability density function (pdf) that was calculated using the parameters estimated from the data. The panels in the middle and on the right show those for Types IV and VI. The distribution shapes of Type I generally include J and U-shaped figures. In this case, we can see the J-shaped distribution. In our preliminary consideration of the STFT series of real speech data, we calculated κ values for each frequency bin shown in Fig. 3. The top, second and third panels show calculated κ values for combined speech data for a female and a male speaker, speech data for a female and speech data for a male, respectively. The distribution of κ is largely the same for the three types of speech data. We see the distribution $(0<\kappa<1)$ of Type IV in the very first frequency range while Type VI for $1<\kappa$ rarely appears. Type I was detected in most frequency bins. This explains why the use of a single super-Gaussian distribution, such as $\tanh$, can perform relatively well in most cases; however, the height and tail of the J-shaped distribution are different for each bin, as shown in the panels in Fig. 2(b). Therefore, a single $\tanh$ assumption may not be sufficiently accurate.

Next, the score function is described by a combination of simple polynomial expressions and distribution parameters that could be described by sample moments [22], [23]. For Pearson Type I, IV, and VI distributions, the original *pdf* and distribution parameters described by using sample moments are summarized in Table I. Expanding the expressions to handle complex values, the score functions for I, IV, and VI are applied such that these distribution parameters indicate

$$\begin{aligned}\text{Type I}: \quad \Phi(\mathrm{Y})&=-\frac{(p+q-2)|Y|-(p+q-2)a-(p-1)b}{|Y|^2-(2a+b)|Y|+a(a+b)}\\ &\quad\times\exp(j\angle Y)\\ \text{Type IV}: \Phi(Y)&=\frac{2b|Y|-2b\mu-2b\tau\delta}{|Y|^2-2\mu|Y|-\mu^2+\tau^2}\exp(j\angle Y)\\ \text{Type VI}: \Phi(\mathrm{Y})&=\frac{(c+1)|Y|+(c+1)a+(\beta-1)\alpha}{|Y|^2-(2a-\alpha)|Y|+a(a-\alpha)}\\ &\quad\times\exp(j\angle Y)\end{aligned} \tag{12}$$

where the distribution parameters of Types I, IV, and VI can be calculated using the formulae shown in the Appendix. Note that b without any suffixes in (12) is different from b_0 and b_1 in (10). As for making the coefficients $\{b_0,b_1,c_0,c_1,c_2\}$ of $|Y|^0$, $|Y|$, and $|Y|^2$ in (10) correspond to the coefficients of $|Y|^0$, $|Y|$, and $|Y|^2$ in (12), we can show that

$$\begin{aligned}\text{Type I}: b_0&=-(p+q-2)a-(p-1)b,\ \ b_1=p+q-2\\ c_0&=a(a+b),\ \ c_1=-(2a+b),\ \ c_2=1\\ \text{Type IV}: b_0&=-2b\mu-2b\delta,\ \ b_1=2b\\ c_0&=-\mu^2+\tau^2,\ \ c_1=-2\mu, c_2=1\\ \text{Type VI}: b_0&=(c+1)a+(\beta-1)\alpha,\ \ b_1=c+1\\ c_0&=a(a-\alpha),\ \ c_1=-2(a-\alpha),\ \ c_2=1.\end{aligned}$$

When applying the Pearson system to frequency-domain BSS, our proposed methods utilize forms (10) and (12) as the score functions. The methods used to estimate the parameters of (10) and (12) are provided in the following sections.

IV. Proposed Methods

With FD-Pearson-ICA, we must estimate the parameters of the score function, defined by (10) or (12). For this, we propose the following three methods.

1) Method 1: Minimization of Cross-Correlation: In this method, we use the score function (10) for learning (4), which is the separation matrix in ICA. To estimate Pearson parameters $\{b_0(f),b_1(f),c_0(f),c_1(f),c_2(f)\}$, we select parameters that minimize the sum of the absolute values of the off-diagonal components of $[\mathbf{I}-\langle\Phi(\mathbf{Y})\mathbf{Y}^H\rangle]$ in (4); that is

$$\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left|\Phi\left(\mathbf{Y}_i(f,m)\right)\mathbf{Y}_j^*(f,m)\right| \tag{13}$$

where $*$ indicates the conjugate. The off-diagonal components represent the higher-order cross-correlation of the outputs. If output signals are well separated, they become mutually independent, and the value of (13) becomes 0. On the other hand, when the separation is incomplete, the absolute value of the off-diagonal components is far from zero. Therefore, we can use off-diagonal components as measures of separation performance. In accordance with this measure, we use a grid search to find the Pearson system parameters $\{b_0(f),b_1(f),c_0(f),c_1(f),c_2(f)\}$ that minimize (13) in an arbitrary range. First, we determined

TABLE I
PEARSON TYPES I, IV, AND VI DISTRIBUTIONS AND PARAMETERS

	Pearson I	Pearson IV
Parameter constraints	$p,q>1,$ $a<x<a+b, b\geq 0$	$0<\tau$
PDF	$\frac{(x-a)^{p-1}}{b^p B(p,q)}\{1-\frac{(x-a)}{b}\}^{q-1}$	$\frac{\Gamma(b+b\delta i)\Gamma(b-b\delta i)\tau^{2b-1}}{\Gamma(b)\Gamma(b+\frac{1}{2})\pi^{\frac{1}{2}}}\frac{\exp\{2b\delta\arctan(\frac{x-\mu}{\tau})\}}{\{(x-\mu)^2+\tau^2\}^b}$
Mean	$a+\frac{bp}{p+q}$	$\frac{b\delta\tau}{b-1}+\mu$
Variance	$\frac{b^2pq}{(p+q+1)(p+q)^2}$	$\frac{\tau^2}{2b-3}\{1+(\frac{b}{b-1}\delta)^2\}$
Skewness	$\frac{2(q-p)\sqrt{p+q+1}}{(p+q+2)\sqrt{pq}}$	$\frac{2b(2b-3)^{\frac{1}{2}}}{(b-1)(b-2)}\frac{\delta}{\{1+(\frac{b}{b-1}\delta)^2\}^{\frac{1}{2}}}$
Kurtosis	$\frac{3(p+q+1)\{2(q^2-pq+p^2)+pq(p+q)\}}{(p+q+3)(p+q+2)pq}$	$\frac{3(2b-3)\{1+\frac{b+2}{b-2}(\frac{b\delta}{b-1})^2\}}{(2b-5)\{1+(\frac{b}{b-1}\delta)^2\}}$

	Pearson VI
Parameter constraints	$\alpha>0, \beta>0, m>0$
PDF	$\frac{\alpha^m(x-a)^{\beta-1}}{B(\beta,m)(\alpha+x-a)^{m+\beta}}$ (Skewness > 0) $\frac{\alpha^m(x-a)^{\beta-1}}{B(\beta,m)(\alpha+a-x)^{m+\beta}}$ (Skewness < 0)
Mean	$a+\frac{\alpha\beta}{m-1}$ (Skewness > 0) $a-\frac{\alpha\beta}{m-1}$ (Skewness <0)
Variance	$\frac{\alpha^2(m+\beta-1)\beta}{(m-1)^2(m-2)}$
Skewness	$\frac{2(2\beta+m-1)\sqrt{m-2}}{(m-3)\sqrt{\beta(\beta+m-1)}}$ (Skewness > 0) $-\frac{2(2\beta+m-1)\sqrt{m-2}}{(m-3)\sqrt{\beta(\beta+m-1)}}$ (Skewness <0)
Kurtosis	$\frac{6[\{\beta^2+(\beta+m-1)^2\}(m-2)+\beta(\beta+m-1)(m-1)]}{(m-4)(m-3)\beta(\beta+m-1)}$

the score functions (10) for the candidates of the parameter sets $\{b_0(f), b_1(f), c_0(f), c_1(f), c_2(f)\}$. For each parameter set, we estimated an unmixing matrix using (4) and obtained separated signals with (3). We compared the off-diagonal component (13) for all unmixing matrices and we select the parameter set that achieves the minimum off-diagonal component. In practice, to avoid the complexity of the parameter grid search, we can express (10) on the Type IV form and freely search the parameter set within the theoretical range shown in Table I.

2) Method 2: Estimation of Appropriate Pearson Distribution Type: This method directly decides the appropriate Pearson type and Pearson parameters for each frequency bin by using (11) and (12). Ideally, in (12), the Pearson parameters based on sample moments should be estimated from a source signal. However, we cannot use source signals in our blind scenario. Therefore, to estimate the sample moments, we propose using pre-separated signals. With this method, we estimate the pre-separation matrix by the previous ICA method and set the matrix as the initial value for FD-Pearson-ICA. As the pre-separation method, we can use any separating method, including Fast-ICA [12], [13] and ICA (4) with conventional $\tanh$ in (6). We label these methods *Method 2-f* and *Method 2-t*, respectively. With the separated signal $\mathbf{Y}(f,m)$ obtained from the initial separation matrix, we calculate sample moments and detect the Pearson type. The concrete calculation procedure in each frequency is organized as follows.

1) Estimate separation matrix $\hat{\mathbf{W}}(f)$ in advance by Fast-ICA or the algorithm using (6) [4], and use it as the initial value $\mathbf{W}_0(f) = \hat{\mathbf{W}}(f)$.
2) Calculate κ (see (11)) by the skewness and kurtosis of the absolute value of $\mathbf{Y}(f,m)$ obtained with (3).
3) Following κ, the appropriate Pearson distribution type is specified, and the parameters of the score function defined in (12) are calculated by the moments of the STFT frame series $|\mathbf{Y}(f,m)|$ according to the Appendix.
4) Renew $\mathbf{W}(f)$ by (4).
5) Iterate procedures 3) and 4) until there is a convergence of (4).

Compared with Method 1, the computational burden is lightened since it is unnecessary to perform a grid search.

3) Method 3: Combining Methods 1 and 2: In Fig. 3, the κ values vary for the lower frequency bins (until 100 frequency bins). On the other hand, at higher frequency (over 100 frequency bins), the κ values are similar among the bins. Moreover, in a preliminary investigation, we found that individual histograms related to the frequencies of the estimated parameters $b_0(f)$, $b_1(f)$, $c_0(f)$, $c_1(f)$, and $c_2(f)$ have similar tendencies for all speaker combinations at higher frequencies. Based on this fact, we propose another method that combines Methods 1 and 2. First, we determine the boundary frequency f_0. Then, the score function in frequency ranges lower than f_0 is always estimated by Method 2, and the fixed pre-estimated average obtained by Method 1 for each parameter of (10) is applied to the score function in frequency ranges higher than f_0. To assure generality, we prepared the averaged parameters obtained by applying Method 1 to a limited number of data combinations. Concretely:

1) calculate the mean values $\{\bar{b}_0(f), \bar{b}_1(f), \bar{c}_0(f), \bar{c}_1(f), \bar{c}_2(f)\}$ of the parameters estimated by applying Method 1 to arbitrary data combinations such as signals of the combined speech of a female and a male, or the combined speech of two females;
2) define f_0 as the boundary point;
3) for low frequencies $f \leq f_0$, apply 3) of Method 2 according to the appropriate Pearson type for each frequency bin based on κ in (11);
4) for high frequencies $f \geq f_0$, input averaged parameters $\{\bar{b}_0(f), \bar{b}_1(f), \bar{c}_0(f), \bar{c}_1(f), \bar{c}_2(f)\}$ for each bin directly into (10).

To choose the best f_0 in advance, we compared the SIR values when using f_0 between 0 and 200 bins and selected the f_0 that provided the highest SIR. The methods that employ *Methods 2-f* and *2-t* for $f < f_0$ are indicated as *Methods 3-f* and *3-t*, respectively.

V. Experimental Results

A. Experimental Conditions

We conducted separation experiments with real speech signals and measured room impulse responses. The speech data were convolved with impulse responses measured in an actual room (Fig. 4) whose reverberation time was 130 ms. As original speech, we used Japanese sentences spoken by male and female speakers. We then made observation signals with (1) and investigated four combinations of speakers. The length of the speech data was 3 s. The STFT frame size was 512, and the frame shift was 256 at a sampling rate of 8 kHz. To solve the permutation problem of frequency-domain ICA, we employed a direction of arrival and correlation approach [10], and to solve the scaling problem we used the minimum distortion principle [25]. For numerical analysis, we arranged four data sets: female and female (*f&f*), two types of female and male (*f&m1*, 2) combinations, and male and male (*m&m*).

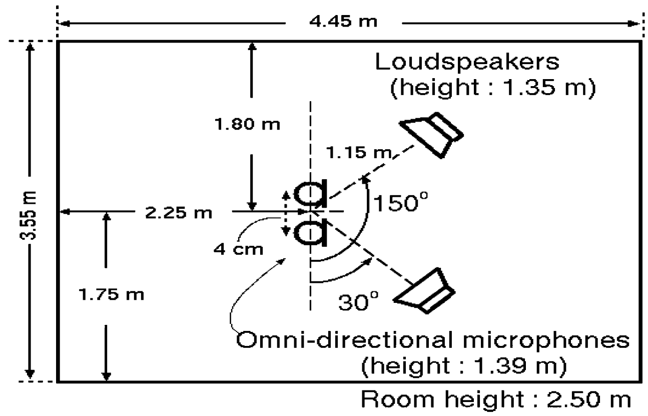

Fig. 4. Room layout used for experiments.

With these methods, we used the signal-to-interference ratio (SIR) as a separation performance measure

$$\mathrm{SIR}_i = 10\log_{10}\frac{\sum_n y_{ii}^2(n)}{\sum_{k\neq i}\left(\sum_n y_{ik}(n)\right)^2} \tag{14}$$

where $y_{ik}(n)$, is a target signal $s_k(n)$-oriented component at $y_i(n)$, that is, $y_i(n) = \sum_{k=1}^{N} y_{ik}(n)$.

To compare the FD-Pearson-ICA methods with other nonlinear functions applied to the score function, we considered conventional $\tanh$ and the GGD-based nonlinear functions [20]. The score function for $\tanh$ is described in (6). For the family of GGD-based nonlinear score functions (7), we searched for the best parameter $\hat{\beta}$ from the range $0.5 \leq \beta < 1.0$ and uniformly defined it in all frequency domains, as in [20].

B. Results

Table II summarizes the results we obtained using *Methods* 1, 2, and 3, conventional $\tanh$, and GGD-based modeling methods for the four types of data sets. With our proposed FD-Pearson-ICA approach, in terms of improved separation performance, we obtained maximum values that were around 3.5 dB better than with conventional $\tanh$ and around 2.5 dB better than with conventional GGD. Although the results vary depending on the combination of speakers, on average our proposed FD-Pearson-ICA achieves better performance than conventional $\tanh$ and GGD.

For conventional nonlinear functions, the GGD-based modeling method was slightly better than $\tanh$. The performance differences are also confirmed in [20].

Method 1 using a grid search worked well for data combinations *f&m2* and *f&f*. The *m&m* combination in *Methods 1* and *2-f* and *f&m2* in *Method 2-f* performed poorly. For these results, we will introduce another criterion to enable us to compare our

TABLE II
SIR (dB) VALUES OBTAINED WHEN EMPLOYING CONVENTIONAL $\tanh$, GGD , AND FIVE FD-PEARSON-ICA METHODS

	tanh	*GGD*
m & m	17.71	17.76 ($\hat{\beta}=0.6$)
f & m 1	15.92	15.89 ($\hat{\beta}=0.6$)
f & m 2	14.99	17.22 ($\hat{\beta}=0.5$)
f & f	17.27	17.76 ($\hat{\beta}=0.7$)
average	16.47	17.16

	Method 1	*Method 2-f*	*Method 2-t*	*Method 3-f*	*Method 3-t*
m & m	17.00	13.68	17.54	17.77	18.12
f & m 1	15.99	17.63	16.38	18.54	16.42
f & m 2	18.61	14.58	17.08	18.09	18.38
f & f	20.12	19.08	18.62	20.24	19.43
average	17.93	16.24	17.41	18.66	18.08

proposed *Method 2-f* with conventional methods. Also, the averaged SIR value in *Method 2-t* indicated better performance than the conventional $\tanh$ and GGD-based approaches.

The separation performance obtained with *Method 3-f* and *Method 3-t* was 14% and 9.6%, respectively, better than that obtained with *Method 2*. Compared to the estimation of whole frequencies, the discrimination of distribution types by κ seems to work particularly well within the lower frequency domain. In this experiment, the averaged parameters used in *Method 3* were estimated by using only two data combinations, *f&m2* and *f&f*, where the separation performance achieved by *Method 1* was better than that for other combinations, and thus these parameters were utilized for all combinations. That is, the parameters averaged by the estimations of two data combinations (*f&m2* and *f&f*) were directly applied to *m&m* and *f&m1*. However, the low SIR value for *m&m* when employing *Method 2-f* was clearly improved by around 4 dB by employing *Method 3-f*. Therefore, the results related to *Method 3* suggest that using parameters pre-estimated by *Method 1* at high frequencies provided better performance, while the parameters estimated with data moments worked well at low frequencies.

C. Discussion

Table II clearly shows that our proposed Methods 3 by FD-Pearson-ICA are better than the above conventional $\tanh$ and GGD. On the other hand, Methods 2 have lower SIR values compared to $\tanh$ and Method 1. Accordingly, we investigated the disparity between the distributions for separated signals obtained with the conventional methods, *Methods 2-f* and *t*, and the true score function. Let nn be the length of the speech signal and $\mathbf{s}(n) = (s_1(n), s_2(n))^T$, $n = 1, \cdots, nn$ as a vector of known signals, while assuming it is obtained from a true score function. Also, let the vector of two separated signals obtained from the parametric score function by $\tanh$, GGD, *Method 2-f* or *Method 2-t* be $\mathbf{y}(n) = (y_1(n), y_2(n))^T$, $n = 1, \cdots, nn$. In our problem, we approximate the distribution by using the signal amplitude histogram. For known signals $s_1(t)$ and $s_2(t)$ and for separated signals $y_1(t)$ and $y_2(t)$, we describe the histograms, and compare the configurations. Fig. 5 shows an example that compares the configuration disparities of the histograms. The number of bins in the histogram was 21. In this case, the white and gray bars show the frequency occurrence for a female speech signal and for a separated signal from the combination *f&m2* by $\tanh$, respectively. As shown by the solid and dotted lines, there were certain differences between the signals. These differences express the distribution disparity between signals obtained by parametric and by true score functions. To investigate the configuration disparities of the histogram, we defined the following measure:

$$B(s_1 : y_1) = \sum_{j=1}^{J} |H_{s_1}(j) - H_{y_1}(j)| \tag{15}$$

where J denotes the total number combining each interval of the histogram, and $H_{s_1}(j)$ and $H_{y_1}(j)$ describe the occurrences at the jth interval in the histograms of s_1 and y_1. Table III summarizes the B-values for $\tanh$, *GGD*, and *Methods 2-f* and *2-t*. The separated signals of channels 1 and 2 obtained from $\tanh$, GGD, and *Methods 2-f* and *2-t* are represented by t_1 and t_2, g_1 and g_2, p_{f1} and p_{f2}, and p_{t_1} and p_{t_2}, respectively. Table III clearly shows that B-values indicate that the distribution of the separated signals by Methods 2 was closer to distribution of true signals. That is, the score functions estimated by *Methods 2-f* and *2-t* were closer to the true value than those estimated by $\tanh$ and GGD. Hence, the B-values show that our approach is superior.

For parameter estimation in Method 1 and Method 3, the accuracy only relies on the optimization procedure. We have estimated the parameters within theoretical range of the Pearson distribution. Concerning the procedure, we can follow the shape for improvement of estimation in each grid, that is, we can confirm whether optimum values are estimated or not, or the local minimum exists or not. Since we considered Method 1 with only grid search procedure to be insufficient for parameter estimation, we proposed Methods 2 and 3 including the procedure that could previously predict the parameters from the distribution type.

Furthermore, we pose two primitive questions: "What performance can we obtain if the shape parameter of GGD is estimated for each frequency?" Moreover, "If we use the averaged parameters calculated by the known impulse responses and separa-

TABLE III
COMPARISON USING B-VALUES FOR $\tanh$, GGD, *Method 2-f* AND *Method 2-t*

		m & m	f & m1	f & m2	f & f	average
tanh	t_1	7114	5889	7114	7430	6443.5
	t_2	5644	7066	5720	6834	6316.0
GGD	g_1	5291	5887	7088	7182	6362.0
	g_2	5626	7058	5386	6886	6239.0
Method 2-f	p_{f1}	3823	4671	5712	5978	5046.0
	p_{f2}	4312	6036	4106	5980	5108.5
Method 2-t	p_{t1}	4245	4831	5826	6014	5229.0
	p_{t2}	4342	5884	4270	5838	5083.5

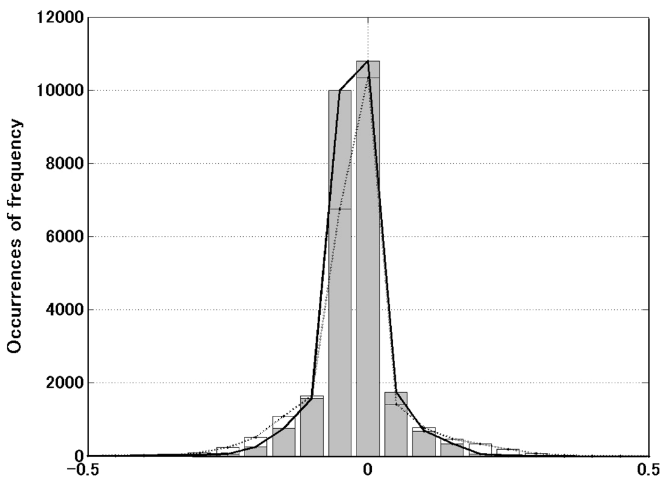

Fig. 5. Histograms for separated signals (white bars: histogram for the original speech signal of a female. The dotted line traces the configuration of the histogram; gray bars: histogram for separated speech signal by $\tanh$. The solid line traces the histogram configuration.)

tion series as the supervised data, we can obtain the optimum of FD-Pearson-ICA. What is the level of its performance?" To deal with these questions, we experimentally examined three more methods, as described below as follows.

1) GGD-ef: Estimation of GGD-Based Score Function for Each Frequency Bin: The shape parameter β in (7) is calculated for each frequency. To consider the optimum usage of the GGD-based score function for each frequency bin, we utilized source signals $s_i(n)$ and room impulse responses $h_{ji}(n)$. This implies supervised *non-blind* source separation. We selected an adequate value β for $0.5 \leq \beta < 1.0$ with which the best SIR was obtained at each frequency. This method is labeled *GGD-ef*.

2) Supervised: Usage of Supervised Impulse Responses: To confirm the best performance of the FD-Pearson-ICA approach, we calculated the SIR under supervised *non-blind* assumptions. As in *Method 4* above, we assume that we know source signals $s_i(n)$ and room impulse responses $h_{ji}(n)$. It should be noted that this is a completely *non-blind* speech separation method. We conducted this experiment to determine the optimum performance with Method 1. With this method, we use the score function defined in (10) and selected the Pearson system parameters $\{b_0(f), b_1(f), c_0(f), c_1(f), c_2(f)\}$, with which the best SIR was obtained at each frequency f by performing a grid search in the appropriate range. This method is labeled *Supervised*.

3) Method 3-s: Method 3 Using Learned Averaged Parameters: We applied the *Supervised* method to two data combinations previously used in Method 3 and averaged parameters. Using the averaged parameters at high frequencies with the *Supervised* method and those at low frequencies with the *Method 2-f* calculation, we conducted separation procedures for all data combinations. This method is labeled *Method 3-s*.

The results for the above three methods are summarized in Table IV. *GGD-ef*, which adopted the score function for each frequency, provided a greater improvement than conventional $\tanh$ or GGD shown in Table II. With the GGD method, estimation of the shape parameter for each frequency bin improved the results over the previous GGD method that applied the estimation uniformly to all frequency bins [20]. This fact suggests that adopting the score function for each frequency bin is an efficient way to improve separation performance. This tendency may apply not only to the GGD method but also to the FD-Pearson-ICA method. Considering this result and the supervised performance, we suggest that an approach that models each data distribution shape is efficient for BSS. In *Supervised* cases with FD-Pearson-ICA, the SIR values indicated the best performance of all the methods. This condition achieved a certain optimum separation performance, and it may further improve if the search range is expanded. Also, we should note that the optimum performance could be estimated by another supervised procedure. In this experiment, *Method 3-s* provided good performance using the learned mean parameters. It should be noted that this method applies a model learned from only two combinations to all combinations. This suggests that learned parameters obtained with a small data set perform well for the open data (for the entire data set). In addition to the method used to calculate the optimum separation performance, we plan to consider how *a priori* knowledge of sources may influence the proposed approach in different ways.

Summarizing, the above results, we found that 1. The optimum separation performance with FD-Pearson-ICA is better than that with GGD; 2. Methods 1 and 3, which are blind, are better than supervised non-blind GGD; 3. The optimum separation performance with Method 3 (*Method 3-s*) was

TABLE IV
SIR (dB) VALUES OBTAINED WHEN EMPLOYING THREE METHODS FOR THE GGD-BASED SCORE FUNCTION ESTIMATED FOR EACH FREQUENCY BIN: (*GGD-ef*), SUPERVISED APPROACH (*Supervised*), AND COMBINED APPROACH (*Method 3-s*) OF *Supervised* AND *Method 3-f*

	GGD-eft	*Supervised*	*Method 3-s*
m & m	18.12	20.78	18.04
f & m 1	15.85	19.62	18.57
f & m 2	17.53	19.34	18.30
f & f	18.10	19.44	20.24
average	17.40	19.80	18.79

near that of *Method 3-f*, that is, Method 3 can be considered effective for separation. Consequently, we believe that the proposed FD-Pearson-ICA is a superior method for solving the frequency-domain BSS problem, although these results were only obtained with four pairs of speakers and the supervised parameters obtained with two pairs of speakers. These are preliminary findings and so we need to conduct more experiments using different pairs of speakers if we are to realize a complete BSS method. Naturally, when the number of speakers involved increases, the computational complexity regarding the learning of the separation matrix equally increases. In our case, since we handle the data within frequency domain, the complexity caused by the number of the sample does not change. In the cases we have proposed FD-Pearson-ICA, we have to estimate the parameters of the Pearson distribution. The number of the parameter is equal to M. Consequently, the computational complexity would increase in terms of $O(M)$ adding to the general complexity by the number of speakers.

Furthermore, we considered the required computational time for performing these methods. We obtained results using a Matlab profile report, which we have summarized in Table V. In this case, the CPU clock speed was 594 MHz. Methods that applied conventional nonlinear functions to the score function were faster than Methods 1 and 2; however, by reducing the optimization procedure, Method 3 could perform at a reasonable computation speed, thus improving performance.

VI. CONCLUSION

To achieve frequency-domain separation matrix estimation with ICA, we proposed a practical parametric Pearson distribution system for the source distribution at each frequency, which could detect the score function. We first confirmed the efficiency of applying the Pearson system to frequency-domain speech BSS under blind conditions with three methods: estimating unknown parameters to minimize the cross-correlation of the separation matrix, directly calculating the transform formulas based on κ discrimination, and a combination of these two methods. The proposed approach significantly improved the separation performance, compared with conventional $\tanh$ and GGD-based modeling approaches. Regarding the parametric score functions of conventional $\tanh$, GGD and FD-Pearson-ICA, the use of a distance measurement showed that FD-Pearson-ICA was closest to the true score function. Through experiments using the proposed FD-Pearson-ICA and GGD-based approaches applied at each frequency, we confirmed that modeling each different distribution shape for each frequency bin is a useful technique as a frequency-domain BSS method. That is, modeling based on the data information was superior as regards separation performance. We have analyzed signal synthesized for real sounds and room impulse response but our approach should also be examined for natural environments.

TABLE V
COMPUTATIONAL TIME FOR PERFORMING CONVENTIONAL NONLINEAR FUNCTIONS AND FD-PEARSON-ICA METHODS

	Applied function	Time [sec]
Conventional nonlinear functions	*tanh*	10.38
	GGD-based	18.09
	GGD-ef	91.94
FD-Pearson-ICA	*Method 1*	6888.66
	Method 2f	52.06
	Method 2t	55.44
	Method 3f	16.70
	Method 3t	17.12
	Supervised	8141.70
	Method 3-s	19.07

APPENDIX

The distribution parameters used in (11) are shown here. These parameters are transferred from the sample moments shown in Table I. A detailed derivation can be seen in [22]–[24].

Pearson Type I:

$$r = \frac{6\left\{\text{Kurt} - (\text{Skew})^2 - 1\right\}}{6 + 3 \times (\text{Skew})^2 - 2 \times \text{Kurt}}$$

$$r_3 = \frac{r}{2} + \frac{r(r+2)}{2}\sqrt{\frac{(\text{Skew})^2}{(\text{Skew})^2(r+2)^2 + 16(r+1)}}$$

$$r_4 = \frac{r}{2} - \frac{r(r+2)}{2}\sqrt{\frac{(\text{Skew})^2}{(\text{Skew})^2(r+2)^2 + 16(r+1)}}$$

$$q = \begin{cases} \max[r_3, r_4] & (\text{Skew} > 0) \\ \min[r_3, r_4] & (\text{Skew} < 0) \end{cases}$$

$$p = \begin{cases} \min[r_3, r_4] & (\text{Skew} > 0) \\ \max[r_3, r_4] & (\text{Skew} < 0) \end{cases}$$

$$b = (p+q)\sqrt{\frac{\text{Var} \times (p+q+1)}{pq}}, \quad a = \text{Mean} - \frac{bp}{p+q}.$$

Pearson Type IV:

$$
\begin{aligned}
b &= \frac{9 + 6 \times (\text{Skew})^2 - 5 \times \text{Kurt}}{6 + 3 \times (\text{Skew})^2 - 2 \times \text{Kurt}} \\
\tau &= \frac{\sqrt{\text{Var} \times \{4(2b-3) - (\text{Skew})^2 (b-2)^2\}}}{2} \\
\delta &= \frac{\sqrt{\text{Var}} \times \text{Skew} \times (b-1)(b-2)}{2b\tau} \\
\mu &= \text{Mean} - \frac{b\delta\tau}{b-1}.
\end{aligned}
$$

Pearson Type VI:

$$
\begin{aligned}
r &= \frac{6\left\{\text{Kurt} - (\text{Skew})^2 - 1\right\}}{6 + 3 \times (\text{Skew})^2 - 2 \times \text{Kurt}} \\
r_1 &= \frac{(r-2)}{2} + \frac{r(r+2)}{2} \sqrt{\frac{(\text{Skew})^2}{(\text{Skew})^2 (r+2)^2 + 16(r+1)}} \\
r_2 &= \frac{(r-2)}{2} - \frac{r(r+2)}{2} \sqrt{\frac{(\text{Skew})^2}{(\text{Skew})^2 (r+2)^2 + 16(r+1)}} \\
\beta &= \max[r_1, r_2] + 1, \quad c = -\min[r_1, r_2] - \beta \\
\alpha &= \sqrt{\frac{\text{Var} \times (c-1)^2 (c-2)}{(c+\beta-1)\beta}} \\
a &= \begin{cases} \text{Mean} - \frac{\alpha\beta}{c-1} & (\text{Skew} > 0) \\ \text{Mean} + \frac{\alpha\beta}{c-1} & (\text{Skew} < 0). \end{cases}
\end{aligned}
$$

References

[1] , S. Haykin, Ed., *Unsupervised Adaptive Filtering*. New York: Wiley, 2000, vol. I, Blind Source Separation.

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

[3] T. W. Lee, *Independent Component Analysis—Theory and Applications*. Norwell, MA: Kluwer, 1998.

[4] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Frequency-domain blind source separation," in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds. New York: Springer, 2005.

[5] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21–34, 1998.

[6] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 320–327, May 2000.

[7] J. Anemüller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation," in *Proc. ICA 2000*, Jun. 2000, pp. 215–220.

[8] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Polar coordinate based nonlinear function for frequency-domain blind source separation," *IEICE Trans. Fundamentals*, vol. E86-A, no. 3, pp. 590–596, 2003.

[9] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 2, pp. 109–116, Mar. 2003.

[10] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 530–538, Sep. 2004.

[11] A. Hyvärinen, "The fixed-point algorithm and maximum likelihood estimation for independent component analysis," *Neural Processing Letters*, vol. 10, no. 1, pp. 1–5, 1999.

[12] A. Hyvärinen, "Fast and robust fixed-point algorithm for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, Mar. 1999.

[13] E. Bingham and A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex value signals," *Int. J. Neural Syst.*, vol. 10, no. 1, pp. 1–8, Feb. 2000.

[14] Z. Koldovsky, P. Tichavsky, and E. Oja, "Efficient variant of algorithm FastICA for independent component analysis attaining the Cramér-Rao lower bound," *IEEE Trans. Neural Netw.*, vol. 17, no. 5, pp. 1265–1277, Sep. 2006.

[15] D. T. Pham and P. Garat, "Blind separation of mixtures of independent sources through a quasi maximum likelihood approach," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1712–1725, Jul. 1996.

[16] K. Pearson, "Memoir on skew variation in homogeneous material," *Philos. Trans. Roy. Soc. A*, vol. 186, pp. 343–414, 1895.

[17] J. Karvanen, J. Eriksson, and V. Koivunen, "Pearson system based method for blind separation," in *Proc. 2nd Int. Workshop ICA and BSS*, 2000, pp. 585–590.

[18] R. Boscolo, H. Pan, and V. P. Roychowdhury, "Independent component analysis based on nonparametric density estimation," *IEEE Trans. Neural Netw.*, vol. 15, no. 1, pp. 55–65, Jan. 2004.

[19] K. Kokkinakis and A. K. Nandi, "Multichannel speech separation using adaptive parameterization of source PDFs," in *Proc. ICA'04, LNCS3195*, C. G. Puntonet and A. Prieto, Eds., 2004, pp. 486–493, Springer-Verlag Berlin, Heidelberg.

[20] R. Prasad, H. Saruwatari, and K. Shikano, "Blind separation of speech by fixed-point ICA with source adaptive negentropy approximation," *IEICE Trans. Fundamentals*, vol. E88-A, no. 7, Jul. 2005.

[21] S. Amari, T. Chen, and A. Cichocki, "Stability analysis of learning algorithms for blind source separation," *Neural Netw.*, vol. 10, no. 8, pp. 1345–1351, 1997.

[22] Y. Nagahara, "The PDF and CF of Pearson type IV distributions and the ML estimation of the parameters," *Stat. Prob. Lett.*, vol. 43, pp. 251–264, 1999.

[23] Y. Nagahara, "Non-Gaussian filter and smoother based on the Pearson distribution system," *J. Time Ser. Anal.*, vol. 24, no. 6, pp. 721–738, 2003.

[24] Y. Nagahara, "A method of simulating multivariate nonnormal distributions by the Pearson distribution system and estimation," *Comp. Statist. Data. Anal.*, vol. 47, no. 1, pp. 1–29, 2004.

[25] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in *Proc. ICA2001*, Dec. 2001, pp. 722–727.

Hiroko Kato Solvang (M'06) received the B.S. degree in physics from Japan Women's University, Tokyo, Japan, in 1989, the M.S. degree in educational physics from Tokyo Gakugei University, Tokyo, in 1988, and the Ph.D. degree in statistical science from the Graduate University for Advanced Studies, Kanagawa, Japan, in 1995, respectively. Her major field of study was time series analysis.

She was employed as a Research Associate at Advanced Telecommunication Research, Information Processing Research Laboratories from 1995 to 1996, and in the Department of Applied Mathematics, Hiroshima University, Hiroshima, Japan, from 1996 to 1998. From 1998 to 2007, she was with NTT Communication Science Laboratories, Kyoto, Japan, as a Research Scientist. Her research at NTT focused on nonlinear/non-Gaussian time series analysis, mathematical statistics, and statistical system analysis related to biomedical and speech signal processing. In 2007 she joined the Department of Genetics at the Norwegian Radium Hospital, where she since has been attached as a Scientist and Project Leader. She is also attached to the Department of Biostatistics, Institute of Basic Medical Science, Oslo University, Oslo, Norway, as a Researcher. Since 2007 she has pursued her statistical research, applying statistical methodologies to expression data, copy number variation, and SNPs array in the research field of genetics related to breast cancer. She has been an Associate Editor of the *Journal of the Japan Statistical Society* since 2007.

Dr. Solvang is a member of the Japan Statistical Society, American Statistical Association, and Institute of Mathematical Statistics.

Yuichi Nagahara received the B.E. degree from Tokyo University, Tokyo, Japan, in 1984, the M.S. degree from Tsukuba University, Tsukuba, Japan, in 1992, and the Ph.D. degree in statistics from Graduated University for Advanced Studies, Hayama, Japan, in 1996.

From 1984 to 1986, he was the Systems Engineer at IBM, Japan. From 1986 to 1997, he was engaged in the research of financial engineering at Nikko Securities. Since 1997, he has been with Meiji University, Tokyo, Japan, and is a Professor in the School of Political Science and Economics. His current research interests include the multivariate non-normal distributions by using the Pearson distributions and its application for various area.

Prof. Nagahara is a member of the Japanese Statistical Society.

Shoko Araki (M'01) received the B.E. and M.E. degrees from the University of Tokyo, Tokyo, Japan, in 1998 and 2000, respectively, and the Ph.D. degree from Hokkaido University, Sapporo, Japan in 2007.

She is with NTT Communication Science Laboratories, NTT Corporation, Kyoto, Japan. Since she joined NTT in 2000, she has been engaged in research on acoustic signal processing, array signal processing, blind source separation (BSS) applied to speech signals, meeting diarization, and auditory scene analysis. She is the author or coauthor of eight book chapters, 17 journal articles, and more than 90 international conference papers.

Dr. Araki is a member of the Organizing Committee of the ICA 2003, the Finance Chair of IWAENC 2003, the co-chair of a special session on undetermined sparse audio source separation in EUSIPCO 2006, and the Registration Chair of WASPAA 2007. She received the 19th Awaya Prize from Acoustical Society of Japan (ASJ) in 2001, the Best Paper Award of the IWAENC in 2003, the TELECOM System Technology Award from the Telecommunications Advancement Foundation in 2004, the Academic Encouraging Prize from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2006, and the Itakura Prize Innovative Young Researcher Award from (ASJ) in 2008. She is a member of the IEICE and ASJ.

Hiroshi Sawada (M'02–SM'04) received the B.E., M.E., and Ph.D. degrees in information science from Kyoto University, Kyoto, Japan, in 1991, 1993, and 2001, respectively.

He joined NTT, Kyoto, Japan, in 1993. He is now a Senior Research Scientist at the NTT Communication Science Laboratories. From 1993 to 2000, he was engaged in research on the computer-aided design of digital systems, logic synthesis, and computer architecture. In 2000, he was with the Computation Structures Group, Massachusetts Institute of Technology, Cambridge, MA, for six months. From 2002 to 2005, he taught a class on computer architecture at Doshisha University, Kyoto. Since 2000, he has been engaged in research on signal processing, microphone array, and blind source separation (BSS). More specifically, he is working on the frequency-domain BSS for acoustic convolutive mixtures using independent component analysis (ICA). He is the author or coauthor of three book chapters, more than 20 journal articles, and more than 80 conference papers.

Dr. Sawada is an Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and a member of the Audio and Electroacoustics Technical Committee of the IEEE Signal Processing Society. He was a tutorial speaker at ICASSP'07. He served as the publications chairs of the WASPAA'07 in New Paltz, NY, and served as an organizing committee member for ICA'03 in Nara, Japan, and the communications chair for IWAENC'03 in Kyoto. He received the Ninth TELECOM System Technology Award for Student from the Telecommunications Advancement Foundation in 1994, and the Best Paper Award of the IEEE Circuit and System Society in 2000. He is a member of the IEICE and ASJ.

Shoji Makino (A'89–M'90–SM'99–F'04) received the B.E., M.E., and Ph.D. degrees from Tohoku University, Sendai, Japan, in 1979, 1981, and 1993, respectively.

He joined NTT Kyoto, Japan, in 1981. He is now a Senior Research Scientist, Supervisor at the NTT Communication Science Laboratories. He was a Guest Professor at the Hokkaido University. His research interests include adaptive filtering technologies and realization of acoustic echo cancellation and blind source separation of convolutive mixtures of speech. He is the author or coauthor of more than 200 articles in journals and conference proceedings and is responsible for more than 150 patents.

Dr. Makino received the ICA Unsupervised Learning Pioneer Award in 2006, the Achievement Award of the IEICE in 1997, the Outstanding Technological Development Award of the ASJ in 1995, the IEEE MLSP Competition Award in 2007, the TELECOM System Technology Award of the TAF in 2004, the Paper Award of the IEICE in 2005 and 2002, the Paper Award of the ASJ in 2005 and 2002, and the Best Paper Award of the IWAENC in 2003. He was a Keynote Speaker at ICA'07 and a Tutorial speaker at ICASSP'07. He is a member of the Award Committee of the IEEE James L. Flanagan Speech and Audio Processing Award. He is a member of the Awards Board and the Conference Board of the IEEE Signal Processing Society (SP). He is an Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and an Associate Editor of the *EURASIP Journal on Applied Signal Processing*. He was a Guest Editor of a Special Issue of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and a Guest Editor of the Special Issue of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-I. He is the Chair of the Technical Committee on Blind Signal Processing of the IEEE CAS Society and a member of the Technical Committee on Audio and Electroacoustics of the IEEE SP Society. He was the Chair of the Technical Committee on Engineering Acoustics of the IEICE and the ASJ. He is a member of the International ICA Steering Committee and a member of the International IWAENC Standing committee. He was the General Chair of the WASPAA'07 in New Paltz, NY, the General Chair of the IWAENC'03 in Kyoto, the Organizing Chair of the ICA'03 in Nara, Japan.