FLEXIBLE SCORE FUNCTIONS FOR BLIND SEPARATION OF SPEECH SIGNALS BASED ON GENERALIZED GAMMA PROBABILITY DENSITY FUNCTIONS

Kostas Kokkinakis and Asoke K. Nandi

Signal Processing and Communications Group,
Department of Electrical Engineering and Electronics,
The University of Liverpool, Brownlow Hill, Liverpool, L69 3GJ, U.K.
{kokkinak, a.nandi}@liv.ac.uk

ABSTRACT

In this contribution, we propose an entirely novel family of flexible score functions for blind source separation (BSS), based on the generalized Gamma family of densities. An efficient maximum likelihood (ML) technique for estimating the parameters of such score functions in an adaptive BSS setup, is also put forward. Simulations indicate that the proposed density model can approximate speech signals more accurately than conventional distributions, which leads to an increase in separation performance and convergence speed.

1. INTRODUCTION

BSS aims to recover a set of unknown signals, the so-called sources from their observed mixtures, based entirely on very little to almost no prior knowledge about the source characteristics or the mixing structure. In its simplest form, the model assumes that the mixtures $\mathbf{x}(t)=[x_1(t),..,x_m(t)]^T\in\mathbb{R}^m$ are in fact linear and instantaneous combinations of the original source signals $\mathbf{s}(t)=[s_1(t),..,s_n(t)]^T \in\mathbb{R}^n$ at each time instant, such that:

$$\mathbf{x}(t) = \boldsymbol{A}\mathbf{s}(t) \tag{1}$$

where $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ denotes the non-singular mixing matrix. In many aspects, BSS is an equivalent process to independent component analysis (ICA), which by definition searches for a linear transformation $\boldsymbol{W}$ that can effectively minimize the statistical dependence between its components [1]. In this context, the recovered source signal estimates in vector form, can be written as:

$$\mathbf{u}(t) = \boldsymbol{W}\mathbf{x}(t) \tag{2}$$

Since it first appeared, the entropy maximization algorithm or Infomax [2], fairly quickly catalyzed a significant surge of interest in using information theory to perform ICA. An efficient way of updating the separation matrix $\boldsymbol{W}$ with respect to its entropy gradient, is the natural gradient algorithm (NGA) [3], which itself is an optimal rescaling of the standard (stochastic) gradient of [2], and is given by:

$$\boldsymbol{W}_{k+1} = \boldsymbol{W}_k + \lambda\big[\boldsymbol{I} - \boldsymbol{\varphi}(\mathbf{u})\mathbf{u}^T\big]\boldsymbol{W}_k \tag{3}$$

where λ denotes the step-size (or learning rate), $\boldsymbol{I}$ is the identity matrix, while vector $\boldsymbol{\varphi}(\mathbf{u}) = [\varphi_1(u_1),\ldots,\varphi_n(u_n)]^T$ represents the nonlinear monotonic activation (or score) functions, described as:

$$\varphi_i(u_i) = -\frac{\partial \log p_{u_i}(u_i)}{\partial u_i} = -\frac{\frac{\partial p_{u_i}(u_i)}{\partial u_i}}{p_{u_i}(u_i)} \tag{4}$$

This work is supported by the Engineering and Physical Sciences Research Council of the U.K. and the University of Liverpool.

Higher-order statistics in (3), are implicitly introduced by choosing suitable nonlinear score functions from (4). In theory, these should be capable of doing an adequate job when modeling the probability density functions (PDFs) of the unknown sources at hand.

2. FLEXIBLE INDEPENDENT COMPONENT ANALYSIS

By design, density matching BSS methods such as Infomax, are predominantly relying on explicit knowledge regarding source signal priors. However, in practical cases when sources exhibiting different densities are mixed together, their capability to distinguish and switch accordingly amongst these, can be significantly compromised. In addition, the separation performance and convergence properties of (3), depend closely on the nature of the nonlinear function $\varphi_i(u_i)$ used to approximate — or to be more precise used to hypothesize upon — the density function of the unknown source signals. Despite the fact that, in some cases, a certain flexibility can be afforded, an ill matched score function can result in a severe model mismatch or a non-converging solution.

Recent contributions have provided some interesting solutions on possible ways to instill source flexibility into BSS schemes (e.g., see [4, 5] and [6]). In this paper, we implement a novel and practically flexible BSS approach, specifically tailored to speech signals. First, we stipulate that the generalized Gamma density (GΓD) model can parameterize the PDFs of the speech sources well. Enough solid evidence supporting such claim, have been documented in [7, 8] and more recently in [9, 10]. Second, we derive an entirely new family of parametric and flexible activation functions, based entirely on the GΓD parent model. It is further shown that commonly used score functions for BSS, are in fact special cases of this new parametric family. Third, potential computational shortcomings when estimating the parameters of the aforementioned activation function, are tackled with an extended ML statistical inference approach, which is based on unconstrained multidimensional optimization [11]. Numerical simulations when an adaptive version of the proposed GΓD-based activation function is used in the NGA update, demonstrate a substantial increase in separation performance and convergence speed, when compared against other more conventional approaches.

3. GENERALIZED GAMMA DENSITY FUNCTION

Traditionally, in the field of signal processing, much work has been focusing on defining accurate mathematical models to characterize the amplitude distribution of a wide class of non-stationary stochas-

1-4244-0469-X/06/$20.00 ©2006 IEEE

tic processes, such as speech (e.g., see [7, 8] and [9, 10, 12]). Towards this direction, a large number of well-known parametric probability density function models, are currently available in the statistical literature. By employing three parameters, the generalized Gamma density (GΓD) model is far more flexible than either the standard Gamma distribution or the so-called generalized Gaussian density (GGD) function [12]. For any zero-mean ($\mu = 0$) signal $x \in \mathbb{R}$, the two-sided GΓD model, as proposed in [13], is equal to:

$$p_x(x|a,\beta,\gamma) = \frac{\gamma\beta^{-a\gamma}}{2\Gamma(a)} |x|^{a\gamma-1} \exp\left[-\left(\frac{|x|}{\beta}\right)^{\gamma}\right] \tag{5}$$

valid for all non-negative values of x. Note that the positive real-valued parameters $a > 0$, $\gamma > 0$ and $\beta > 0$, collectively define the shape and the scale of the density function, respectively, while also in (5), $\Gamma(\cdot)$ denotes the complete Gamma function, which in turn is given by:

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx, \quad z > 0 \tag{6}$$

Special cases of the GΓD include well-known two parameter distributions, namely the GGD ($a\gamma = 1$) and the Gamma density ($\gamma = 1$), as well as standard single parameter distributions, for example the Laplacian density ($a = 1, \gamma = 1$) and the Gaussian (or normal) distribution ($a = 0.5, \gamma = 2$).

4. NOVEL FLEXIBLE SCORE FUNCTIONS

In an effort to meet the need of accurately approximating the PDFs of unknown speech sources in the context of BSS, we propose an entirely novel family of parametric (or flexible) activation functions. After substituting (5) into (4) for the source signal estimates, the derived score function can be seen to inherit a nice generalized parametric structure, which in turn can be attributed to the highly flexible GΓD parent model. In such case, after some simple calculus, the proposed score function[1] can be written as:

$$\varphi(u_i|a,\beta,\gamma) = \frac{\text{sign}(u_i)}{|u_i|}\left(\frac{\gamma}{\beta^{\gamma}}|u_i|^{\gamma} - a\gamma + 1\right) \tag{7}$$

which is valid for all $u_i > 0$. In principle, $\varphi(u_i|a,\beta,\gamma)$ is able to sufficiently model a large number of speech signals, as well as several other types of heavy- and light-tailed distributions, since its characterization depends explicitly on all three parameters, a, β and γ. Hence, commonly used parametric and flexible activation functions can be obtained simply by substituting appropriate values for parameters a, β and γ in (7). For instance, a scaled form of the GGD-based score function (see [4, 6] for its normalized form), constitutes such a special case of (7), when $a\gamma = 1$ and $\beta = 1$:

$$\varphi(u_i|\gamma) = \gamma\,\text{sign}(u_i)\,|u_i|^{\gamma-1} \tag{8}$$

which is valid for all $\gamma \geq 1$, while another special case is the very popular $\varphi(u_i) = \text{sign}(u_i)$ suitable for sources exhibiting a Laplacian PDF. The proposed family of the GΓD-based parametric score functions are depicted in Fig. 1, where they are plotted for different values of the shape parameters a and γ. Note that in some special cases, essentially those corresponding to heavy-tailed (or sparse) distributions, function $\varphi(u_i|a,\beta,\gamma)$ could become singular for $u_i = 0$.

[1]Note that throughout the derivation of function $\varphi(u_i|a,\beta,\gamma)$, which is here omitted due to lack of space, we have also used the transformation $\text{sign}(u_i) = u_i/|u_i|$, for $u_i \neq 0$.

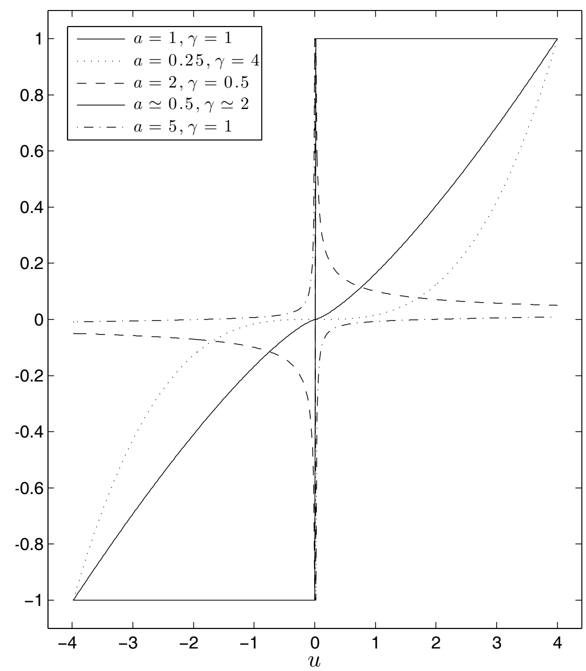

Fig. 1. GΓD-based flexible score functions arising from (7), when plotted for different values of the shape parameters a and γ. Solid lines indicate single parameter functions and dashed lines, two parameter distributions. Note that in all cases, $\beta = 1$.

In practice, to circumvent such problem, the denominator of (7) can be modified slightly to read:

$$\varphi(u_i|a,\beta,\gamma) = \frac{\text{sign}(u_i)}{[|u_i| + \epsilon]}\left(\frac{\gamma}{\beta^{\gamma}}|u_i|^{\gamma} - a\gamma + 1\right) \tag{9}$$

where ϵ is a very small positive parameter (typically around 10^{-4}) which, when put to use, can almost always guarantee that the singularity of (7) for values around $u_i = 0$, is avoided.

5. GENERALIZED GAMMA PDF PARAMETER ESTIMATION

Historically, the development of inference procedures for the GΓD has been a fairly complicated task [10, 14]. Moment matching estimators (MMEs) and maximum likelihood estimators (MLEs), are standard tools for statistical inference. MMEs are simple to deduce, yet these are often susceptible to large estimation errors, whilst MLEs are more efficient, but less convenient to derive and calculate from the data. The inference technique we present here, is a hybrid of the aforementioned approaches. First, we use MME to calculate an initial guess of the shape parameters, which is then refined further by resorting to the MLE. The qth-order absolute central moment of the generalized Gamma density function can be defined as:

$$\mathrm{E}\left[|X|^q\right] = \int_{-\infty}^{+\infty} |x|^q \, p_x(x|a,\beta,\gamma)\, dx \tag{10}$$

where $\mathrm{E}[\,\cdot\,]$ represents the expectation operator. Substituting (5) into (10), it is possible to show that the qth-order central moment transform of the two-sided GΓD model is equal to:

$$m_q = \mathrm{E}\left[|X|^q\right] = \beta^q \left(\frac{\Gamma(a + q/\gamma)}{\Gamma(a)}\right), \ \forall\ q \geq 0 \tag{11}$$

Adopting the notation above, we may further define moment ratios as follows:

$$\mathcal{M}_1(a,\gamma) \;=\; \frac{m_2}{m_1^2} \;=\; \frac{\Gamma(a+2/\gamma)\,\Gamma(a)}{\Gamma^2(a+1/\gamma)} \tag{12}$$

$$\mathcal{M}_2(a,\gamma) \;=\; \frac{m_4}{m_2^2} \;=\; \frac{\Gamma(a+4/\gamma)\,\Gamma(a)}{\Gamma^2(a+2/\gamma)} \tag{13}$$

where the scale parameter β originally present in (11), is being effectively eliminated in both (12) and (13). Based on matching the moments of the data with those of the GΓD, the simultaneous solution of (12) and (13) yields the initial moment estimates for the shape parameters $\hat{a}$ and $\hat{\gamma}$. At this stage the scale parameter is set to $\hat{\beta} = 1$. To refine those further, we resort to ML. In general, for a sequence of mutually independent data $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ of sample size N with density $p_{x_i}(x_i|a,\beta,\gamma)$, the ML estimates are uniquely defined by their log-likelihood function [10, 15]:

$$L(\mathbf{x}|a,\beta,\gamma) = \log\prod_{i=1}^{N} p_{x_i}(x_i|a,\beta,\gamma) \tag{14}$$

$$= N\log\frac{\gamma\beta^{-a\gamma}}{2\Gamma\,(a)} - \frac{1}{\beta^{\gamma}}\sum_{i=1}^{N}|x_i|^{\gamma} + (a\gamma-1)\sum_{i=1}^{N}\log|x_i| \tag{15}$$

Normally, ML estimates are obtained by first differentiating the log-likelihood function in (15) with respect to the GΓD parameters a, β and γ and by then equating those derivatives to zero (e.g., see [10]). Nonetheless, the numerical calculations involved with such an approach are often prohibitive. Instead, here we choose to minimize the ML equation in (15) by exploiting the Nelder-Mead (NM) optimization method [11]. The NM simplex optimization technique, is an enormously popular direct search method for multidimensional unconstrained nonlinear minimization. Its huge appeal in this case, lies in the fact that it can minimize the scalar-valued ML objective function in (15) using function values only, essentially without the need to resort to any derivative information (explicit or implicit). Minimizing with the NM technique to produce the refined ML shape estimates $\hat{a}$ and $\hat{\gamma}$, is computationally efficient. Following this, an estimate for the scale GΓD parameter $\hat{\beta}$ can be also calculated as shown in [15]:

$$\hat{\beta} = m_1\,\frac{\Gamma(\hat{a})}{\Gamma(\hat{a}+1/\hat{\gamma})} \tag{16}$$

whereby it is stipulated that the shape parameters $\hat{a}$ and $\hat{\gamma}$ are known.

6. EXPERIMENTAL RESULTS

First, the proposed inference technique is gauged for a relatively small number of samples ($N = 1000$). For such a sample size, we generate 100 different zero-mean and unit-variance i.i.d. sequences, each for data selected from the most commonly used distributions, often chosen to model speech, namely Gaussian, Laplacian and also Gamma densities, which here we attempt to fully characterize by employing the GΓD model. Table 1 shows the mean calculated for the shape parameters a and γ and the scale parameter β, for every density, after averaging over the same 100 Monte Carlo runs. As evidenced from the results, the proposed extended maximum likelihood inference technique performs exceptionally well. Having established an accurate method for estimating the parameters of the GΓD model, emphasis is now shifted to BSS. In particular, focusing entirely on speech signals, we aim to show that by directly estimating the PDFs of the sources at hand, through an adaptive tuning of the GΓD-based parametric activation function, will almost always lead to significant increase in separation performance. The source material used in these numerical simulations, is taken from the TIMIT speech corpus [16] and it consists of two pairs of male and female 2 s speech signals, sampled at 8 kHz.

Distribution	a	γ	β
Gaussian	0.501	2.032	1.156
Laplacian	1.018	1.002	1.001
Gamma	0.503	1.004	1.087

Table 1. GΓD shape (a, γ) and scale (β) parameter estimates for some typical densities, based on the inferential procedure outlined in Section 5. Sample size is $N = 1000$. Results are averaged 100 independent Monte Carlo runs.

In order to adequately capture the highly non-stationary characteristics of the speech sources, we resort to a block-based batch implementation of the NGA update shown in (3). The mixing matrix $\boldsymbol{A}$ is chosen to be random and fixed, while the update is carried out using short-time blocks of around 150 ms (1200 samples). Three different approaches, with three different activation functions are studied. In all cases, the learning rate parameter λ is tuned for maximum performance, while all algorithms are initialized according to $\boldsymbol{W}(0) = 0.1\boldsymbol{I}$. First, the speech mixtures are passed through the so-called extended Infomax approach, which uses a simple fixed (switching) nonlinearity as described in [5]. Next, the GGD-based score function suggested in [4] and shown here in (8), is put to use. In this case, the score function is adapted using (previous) estimates of the source data. Note that here the adaptation of the exponent parameter for each of the two signals, can be carried out by resorting to one of the moment matching estimators analyzed extensively in [12]. Finally, the NGA update is coupled with the newly proposed GΓD-based activation function given by (9). This is also implemented in a continuously adaptive fashion. However, here the shape and scale parameters of the GΓD are estimated in a near closed-form manner, as outlined in Section 5. The performance of the algorithms during adaptation, is monitored using the cross-talk error metric, as suggested in [3]:

$$\rho = \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{|p_{ij}|^2}{\max\limits_{1\le\ell\le n}|p_{i\ell}|^2} - 1\right) + \sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{|p_{ij}|}{\max\limits_{1\le\ell\le n}|p_{\ell j}|} - 1\right) \tag{17}$$

which is valid for all $i \neq j$. Note that here, p_{ij} define the elements of the permutation matrix $\boldsymbol{P} = \boldsymbol{WA}$, which after assuming that all the sources have been successfully separated, ideally reduces to a permutated and scaled version of the identity matrix.

Fig. 2 depicts the evolution of the performance index $\rho(k)$ defined above, when plotted against the number of iterations, which here is set equal to the total number of samples. As revealed, the proposed GΓD-based flexible ICA approach, considerably outperforms the extended Infomax approach [5]. The proposed method, also manages to achieve an extremely satisfactory steady-state value for ρ, which settles to approximately -30 dB. In addition, it converges faster to the correct solution, while in general requires about $3\times$ fewer iterations than the GGD-based driven NGA update and around $8\times$ fewer iterations if compared against the extended ICA approach, which operates on a fixed activation function regime. Worth noting is also that the difference in performance between the GΓD and the

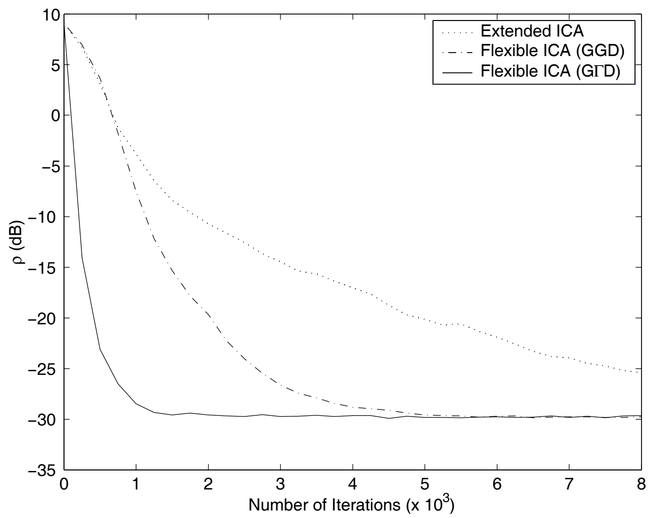

Fig. 2. Evolution of the average value (dB) of the performance metric $\rho(k)$ for the extended and the GGD- and GΓD-based flexible ICA approaches.

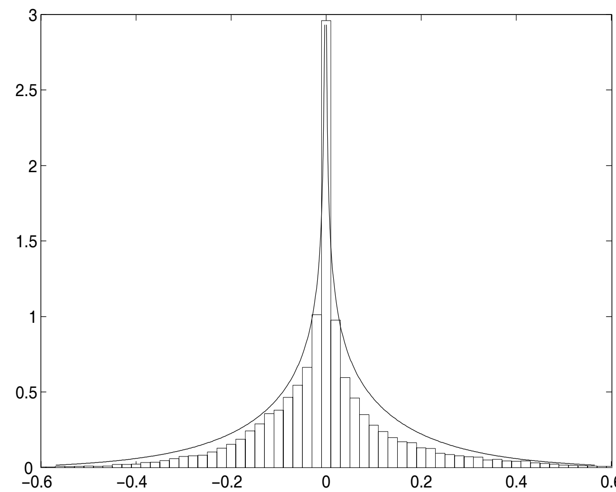

Fig. 3. PDF of the GΓD model for $a = 0.717$, $\gamma = 1.059$ and $\beta = 1.014$, when fitted against the histogram of the actual distribution of the recovered speech source estimate u_1.

GGD-based methods after a solution has been reached, is found to be negligible. This can be attributed to the fact that in the particular experimental setup both activation functions handle and adapt to the speech data equally well. Nonetheless, the GΓD-based score function is still expected to generally favour speech sources with more challenging (or sparse) distributions. Such an argument can be strengthened even further by observing Fig. 3, which essentially illustrates how the GΓD model fits against the actual PDF of the first male speech source signal estimate. To carry out the model fitting, the recovered speech signal is first processed as a series of independent short-time frames. The frame length is set to 150 ms, while the overlapping between successive frames is equal to 50%. The shape and scale parameter estimates a, γ and β are then calculated by resorting to the hybrid ML inference procedure and by averaging across all their individual values at each frame. In this case, the values obtained for the GΓD model are $a = 0.717$, $\gamma = 1.059$ and $\beta = 1.014$. Note that for the same speech source, the GGD model of (8), has yielded $\gamma = 0.581$ as the optimal value for the exponent.

7. CONCLUSIONS

In this paper, we have derived a novel parametric family of flexible activation functions, based exclusively on the GΓD model. To calculate the parameters of these functions in an adaptive BSS setup, we have chosen to minimize the ML equation with the NM simplex optimization method. In theory, this should alleviate excessive computational cost requirements and allow for a fast practical implementation of the proposed technique. Motivated by the widespread appeal of the entropy maximization algorithm, the newly proposed family of score functions has also been applied to linear instantaneous mixtures of speech signals. Experimental results demonstrate that when coupled with the NGA, the GΓD-based score yields a consistent increase in convergence speed and separation performance, and thus appears to be a very promising alternative over existing functions.

8. REFERENCES

[1] J.-F. Cardoso, "Blind Signal Separation: Statistical Principles," *Proc. IEEE*, Vol. 86, No. 10, pp. 2009–2025, Oct. 1998.

[2] A. J. Bell and T. J. Sejnowski, "An Information Maximization Approach to Blind Separation and Blind Deconvolution," *Neural Computat.*, Vol. 7, No. 6, pp. 1129–1159, Jul. 1995.

[3] S.-I. Amari, A. Cichocki and H. H. Yang, "A New Learning Algorithm for Blind Signal Separation," In *Adv. in Neural Informat. Process. Systems*, MA: MIT Press, 1996, Vol. 8, pp. 757–763.

[4] S. Choi, A. Cichocki and S.-I. Amari, "Flexible Independent Component Analysis," *J. VLSI Signal Process.*, Vol. 26, No. 1, pp. 25–38, Aug. 2000.

[5] T.-W. Lee, M. Girolami and T. J. Sejnowski, "Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Sub-Gaussian and Super-Gaussian Sources," *Neural Computat.*, Vol. 11, No. 2, pp. 417–441, Feb. 1999.

[6] K. Kokkinakis and A. K. Nandi, "Multichannel Blind Deconvolution for Source Separation in Convolutive Mixtures of Speech," *IEEE Trans. on Audio, Speech, and Language Process.*, Vol. 14, No. 1, pp. 200–212, Jan. 2006.

[7] W. B. Davenport, "An Experimental Study of Speech-Wave Probability Distributions," *J. Acoust. Soc. America*, Vol. 24, No. 4, pp. 390–399, Jul. 1952.

[8] D. L. Richards, "Statistical Properties of Speech Signals," *Proc. of the IEE*, Vol. 111, No. 5, pp. 941–949, May 1964.

[9] S. Gazor and W. Zhang, "Speech Probability Distribution," *IEEE Signal Process. Lett.*, Vol. 10, No. 7, pp. 204–207, Jul. 2003.

[10] J.W. Shin, J.-H Chang and N. S Kim,"Statistical Modeling of Speech Signals Based on Generalized Gamma Distribution," *IEEE Signal Process. Lett.*, Vol. 12, No. 3, pp. 258–261, Mar. 2005.

[11] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," *Comput. J.*, Vol. 7, No. 4, pp. 308–313, Jan. 1965.

[12] K. Kokkinakis and A. K. Nandi, "Exponent Parameter Estimation for Generalized Gaussian Probability Density Functions with Application to Speech Modeling," *Signal Process.*, Vol. 85, No. 9, pp. 1852–1858, Sep. 2005.

[13] E. W. Stacy, "A Generalization of the Gamma Distribution," *Ann. Math. Statist.*, Vol. 33, No. 3, pp. 1187–1192, Sep. 1962.

[14] J. F. Lawless, " Inference in the Generalized Gamma and Log Gamma Distributions," *Technometr.*, Vol. 22, No. 3, pp. 409–419, Aug. 1980.

[15] A. C. Cohen and B. J. Whitten, *Parameter Estimation in Reliability and Life Span Models*, New York: Marcel Dekker, 1988.

[16] W. Fisher, G. Doddington and K. Goudie-Marshall, "The DARPA Speech Recognition Research Database: Specifications and Status," In *Proc. of the DARPA Speech Recogn. Work.*, pp. 93-99, 1989.