MAP Estimators for Speech Enhancement Under Normal and Rayleigh Inverse Gaussian Distributions

Richard C. Hendriks and Rainer Martin, *Senior Member, IEEE*

***Abstract*—This paper presents a new class of estimators for speech enhancement in the discrete Fourier transform (DFT) domain, where we consider a multidimensional normal inverse Gaussian (MNIG) distribution for the speech DFT coefficients. The MNIG distribution can model a wide range of processes, from heavy-tailed to less heavy-tailed processes. Under the MNIG distribution complex DFT and amplitude estimators are derived. In contrast to other estimators, the suppression characteristics of the MNIG-based estimators can be adapted online to the underlying distribution of the speech DFT coefficients. Compared to noise suppression algorithms based on preselected super-Gaussian distributions, the MNIG-based complex DFT and amplitude estimators lead to a performance improvement in terms of segmental signal-to-noise ratio (SNR) in the order of 0.3 to 0.6 dB and 0.2 to 0.6 dB, respectively.**

***Index Terms*—Maximum *a posteriori* (MAP) estimation, multidimensional normal inverse Gaussian (MNIG) distribution, speech enhancement.**

I. Introduction

THE recent increased use of mobile speech processing applications, like cellular phones and hearing aids, led to an increased interest for speech enhancement algorithms as well. Often, mobile speech processors are used in an environment with a high level of ambient noise, leading to degraded speech quality. Speech enhancement methods, like single-microphone algorithms, can be used as a preprocessor to reduce the noise level before being further processed.

Single-microphone speech enhancement algorithms are often implemented in the spectral domain, e.g., using the discrete Fourier transform (DFT). Clean speech DFT coefficients are estimated by applying a gain function to the noisy DFT coefficients. The gain function is often derived by assuming a certain distribution for the speech and noise DFT coefficients while optimizing for a certain criterium like minimum mean square error (MMSE) or maximum *a posteriori* (MAP) applied on the complex or magnitude DFT coefficients.

Often, those distributions are assumed to be Gaussian, which is supported by the central limit theorem as each DFT coefficient is computed as a sum of time samples. However, the central limit theorem is not applicable under the relatively short frame sizes used in practice. Several studies indicated that super-Gaussian distributions that are more heavy-tailed, like the Laplace or Gamma distribution, give a better description of the distribution of speech DFT coefficients. Therefore, more recently, clean speech estimators have been derived under super-Gaussian distributions. In [1]–[3], MMSE estimators for clean speech DFT coefficients were presented under both a Gamma and a Laplace distribution. In [4], MAP amplitude estimators were presented under an approximation of super-Gaussian densities by parametric probability density functions.

Although super-Gaussian distributions give clearly a better description of the speech DFT distribution than a Gaussian distribution, the shape of the distributions in the aforementioned methods is always chosen fixed over time and frequency. It could be advantageous to derive clean speech estimators under a distribution that can be adapted over time and frequency. At first, in [5], [6], it was experimentally shown that the type of distribution that is most appropriate is speech sound and frame size dependent. Second, the aforementioned methods all implicitly assume the clean speech variance to be known and typically apply the decision-directed approach [7] for speech variance estimation. However, speech data in a certain frame is to some degree nonstationary due to the fact that the variance is not entirely fixed. Such a variation on the variance gives rise to variations in the shape of the observed distribution.

In this paper, we present a new class of speech estimators that take variations of the speech distributions over time and frequency into account and incorporate the uncertainty on the speech variance. More specifically, we present a class of clean speech estimators based on scale mixtures of normals [8], [9]. Scale mixtures of the normal distribution assume that the variance is not fixed, but take the distribution of the variance into account. In [10], Barndorff–Nielsen presented the normal inverse Gaussian distribution (NIG) to model stochastic volatility of heavy-tailed data for financial data modelling applications. More recently, the NIG distribution and its multivariate extension, known as the multidimensional NIG (MNIG) [11], have shown to be very suitable to model a large class of heavy-tailed processes [11]. We assume that speech DFT coefficients follow an MNIG distribution, motivated by the fact that speech is a heavy-tailed process as well. Under this assumption, the speech DFT amplitudes can be shown to be Rayleigh inverse Gaussian (RIG) distributed. Under the MNIG and RIG distribution, we derive clean speech MAP estimators for the complex DFT coefficients and speech amplitudes, respectively. The MNIG and RIG distribution parameters of the resulting gain functions are

Manuscript received February 15, 2006; revised September 9, 2006. This work was supported in part by Philips Research and the Technology Foundation STW, STW, applied science division of NWO and the technology programme of the Ministry of Economics Affairs.. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Israel Cohen.

R. C. Hendriks is with the Department of Mediamatics of Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: r.c.hendriks@tudelft.nl).

R. Martin is with the Institute of Communication Acoustics Ruhr-Universität Bochum, 44780 Bochum, Germany (e-mail: rainer.martin@rub.de).

Digital Object Identifier 10.1109/TASL.2006.889753

adapted to the speech signal using the expectation-maximization procedure presented in [11]. As such, the suppression characteristics are adapted to the observed distribution of the speech DFT coefficients.

The remaining sections of this paper are organized as follows. In Section II, the MNIG distribution and its most relevant properties are presented. In Section III, we present the assumed speech model and derive MAP estimators for complex DFT coefficients and amplitudes. Further, in Section IV, experimental results are presented. Finally, in Section V, conclusions are drawn.

II. Normal and Rayleigh Inverse Gaussian Distribution

In order to facilitate our discussion on the clean speech estimators, we summarize in this section the relevant properties of the MNIG and derive the RIG distribution. For more detailed information, we refer the reader to [10], [11]. Our notation in this article is such that capitals and small symbols indicate random variables and realizations, respectively. Bold symbols indicate the use of matrices.

A d-dimensional MNIG distributed random variable $\mathbf{X}$ is defined as

$$\mathbf{X} = \boldsymbol{\mu} + \Lambda_X \boldsymbol{\Gamma}\boldsymbol{\beta} + \sqrt{\Lambda_X}\boldsymbol{\Gamma}^{\frac{1}{2}}\mathbf{Z} \tag{1}$$

where $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{I})$ and where $\mathbf{X}$ given Λ_X has a Gaussian distribution, i.e., $\mathbf{X}|\Lambda_X \sim N_d(\mu + \Lambda_X\boldsymbol{\Gamma}\beta, \Lambda_X\boldsymbol{\Gamma})$, with $\Lambda_X \sim IG(\delta^2, \alpha^2 - \boldsymbol{\beta}^{\mathbf{T}}\boldsymbol{\Gamma}\boldsymbol{\beta})$. IG denotes the inverse Gaussian distribution with scalar parameters $\alpha > 0$ and $\delta > 0$, vector parameters $\boldsymbol{\beta} \in \mathbb{R}^d$ and $\boldsymbol{\mu} \in \mathbb{R}^d$, and a correlation matrix $\boldsymbol{\Gamma} \in \mathbb{R}^{d\times d}$, which is assumed to be positive definite. The IG distribution is defined for $\lambda_X > 0$ as

$$f_{\Lambda_X}(\lambda_X) = \left(\frac{\delta^2}{2\pi\lambda_X^3}\right)^{1/2} \exp\left[\sqrt{\delta^2(\alpha^2 - \boldsymbol{\beta}^{\mathbf{T}}\boldsymbol{\Gamma}\boldsymbol{\beta})}\right] \times \exp\left[-\frac{1}{2}\left(\delta^2\lambda_X^{-1} + (\alpha^2 - \boldsymbol{\beta}^{\mathbf{T}}\boldsymbol{\Gamma}\boldsymbol{\beta})\lambda_X\right)\right]. \tag{2}$$

The name inverse Gaussian was introduced by Tweedie [12] who noted the inverse relationship between cumulant generating functions of IG distributions and those of Gaussian distributions.

The MNIG distribution of $\mathbf{X}$ is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \int f_{\mathbf{X}|\Lambda_X}(\mathbf{x}|\lambda_X) f_{\Lambda_X}(\lambda_X) d\lambda_X \tag{3}$$

$$= \frac{\delta}{2^{(d-1)/2}} \left(\frac{\alpha}{\pi q(\mathbf{x})}\right)^{(d+1)/2} \times \exp\left[p(\mathbf{x})\right] K_{(d+1)/2}\left[\alpha q(\mathbf{x})\right] \tag{4}$$

with

$$p(\mathbf{x}) = \delta\sqrt{\alpha^2 - \boldsymbol{\beta}^T\boldsymbol{\Gamma}\boldsymbol{\beta}} + \boldsymbol{\beta}^T(\mathbf{x} - \boldsymbol{\mu}) \tag{5}$$

and

$$q(\mathbf{x}) = \sqrt{\delta^2 + \left[(\mathbf{x} - \boldsymbol{\mu})^T\boldsymbol{\Gamma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right]} \tag{6}$$

where $K_{(d+1)/2}$ denotes the modified Bessel function of the second kind with order $(d+1)/2$. Notice that when d is even, $K_{(d+1)/2}$ can be written with a closed-form expression, which enables computationally more efficient implementations.

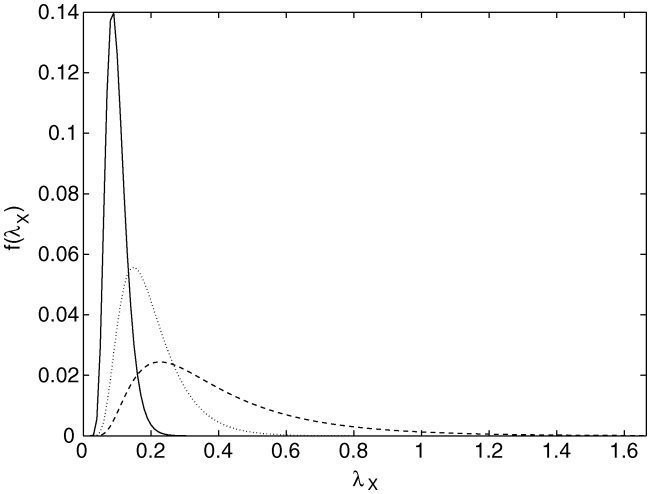


Fig. 1. IG distribution for $(\delta, \alpha) = (1, 10)$ (solid), $(\delta, \alpha) = (1, 5)$ (dotted), and $(\delta, \alpha) = (1, 2.5)$ (dashed).

The density $f_{\mathbf{X}}(\mathbf{x})$ is parameterized by α, $\boldsymbol{\beta}$, δ, and $\boldsymbol{\mu}$. The shape of the density is determined by α such that the smaller α is, the heavier the tails become. Parameter $\boldsymbol{\beta}$ determines the skewness of the density, which means that for $\boldsymbol{\beta} \neq 0$, the density will be asymmetrical. Further, δ is the scale parameter and $\boldsymbol{\mu}$ a translation parameter. Using the MNIG distribution to model the clean speech, DFT coefficients leads to a very flexible framework where the distribution can be adapted to the speech signal by estimation of the parameters, e.g., using the expectation-maximization algorithm as presented in [11]. In this paper, we assume the distribution of $\mathbf{X}$ is symmetrical with zero mean, this means that $\boldsymbol{\mu} = 0$ and $\boldsymbol{\beta} = 0$. Those choices for $\boldsymbol{\mu}$ and $\boldsymbol{\beta}$ will be used in the remainder of this paper.

Although the MNIG probability density function (3) appears to be rather complex, its cumulant generating function has a relatively simple form, that is

$$\Psi_{\mathbf{X}}(\boldsymbol{\omega}) = \delta\left[\alpha - \sqrt{\alpha^2 - (j\boldsymbol{\omega})^T\boldsymbol{\Gamma}(j\boldsymbol{\omega})}\right]. \tag{7}$$

From the cumulant generating function, it follows that the Gaussian distribution is a limiting distribution of $f_{\mathbf{X}}(\mathbf{x})$ when $\alpha \to \infty$, what also becomes clear from (3) when observing the shape of the IG distribution for increasing α in Fig. 1. In Fig. 1, it is shown for $\delta = 1$ that when α becomes larger, the IG distribution becomes more and more peaked and will become a delta impulse for $\alpha \to \infty$. Therefore, (3) becomes in the limit equivalent to a Gaussian distribution. Furthermore, from the cumulant generating function, it follows that the covariance matrix of the MNIG distribution is given by [10], [11] (with $\boldsymbol{\mu} = 0$ and $\boldsymbol{\beta} = 0$)

$$\boldsymbol{\Sigma} = \frac{\delta}{\alpha}\boldsymbol{\Gamma}. \tag{8}$$

In Fig. 2, we show some examples of an NIG probability density function $f_{\mathbf{X}}(\mathbf{x})$ for $\boldsymbol{\beta} = \boldsymbol{\mu} = 0$, $E[X^2] = 1$, and several combinations of δ and α. Comparing this with a zero-mean

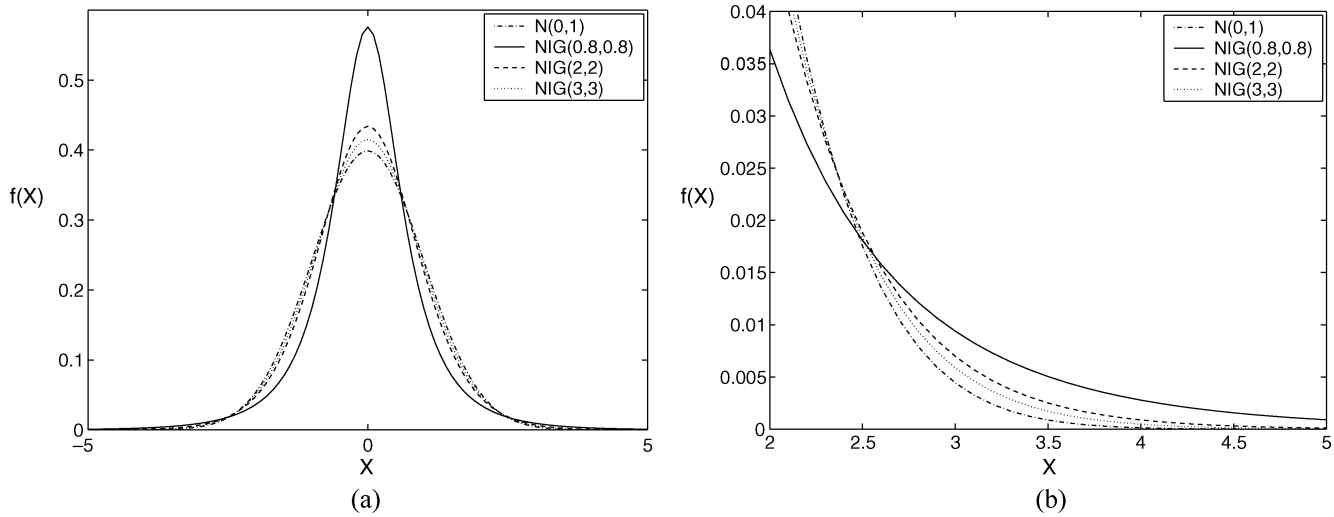


Fig. 2. Normal inverse Gaussian distribution for several values of δ and α, and the Gaussian distribution, all with $E[X^2] = 1$. Notice that the NIG converges towards a Gaussian distribution as α increases. The abbreviations NIG and N in the legend indicate the normal inverse Gaussian and the Gaussian distribution, respectively.

Gaussian distribution, we see that the NIG distribution approximates the Gaussian distribution as α gets larger. Further, the NIG distributions become more peaked and heavy-tailed as α becomes smaller.

The RIG distribution can be derived from the 2-D MNIG distribution by a transformation of (3) into polar coordinates [13]. Consider a 2-D vector $\mathbf{X}$, e.g., $\mathbf{X} = [X_R, X_I]$, with $\mathbf{X} \sim MNIG(\delta, \alpha, \boldsymbol{\mu}, \boldsymbol{\beta}) = MNIG(\delta, \alpha, 0, 0)$ and $\boldsymbol{\Gamma} = \mathbf{I}$. Then, the distribution of the amplitude $A = \sqrt{X_R^2 + X_I^2}$ of $\mathbf{X}$ can be shown to be a scale mixture of Rayleigh distributions.

Consider therefore the 2-D case of (3), that is

$$f_{\mathbf{X}}(x_R, x_I) = \frac{\delta}{\sqrt{2}} \left(\frac{\alpha}{\pi \sqrt{\delta^2 + x_R^2 + x_I^2}} \right)^{3/2} \exp[\delta\alpha] K_{3/2} \left[\alpha \sqrt{\delta^2 + x_R^2 + x_I^2} \right]. \quad (9)$$

Transformation of (3) into polar coordinates with $X_R = A\cos(\Phi)$ and $X_I = A\sin(\Phi)$, Jacobian A and integration over Φ then gives

$$f_A(a) = \int_0^{2\pi} f_{A,\Phi}(a, \phi) d\phi = \frac{a\sqrt{2}\alpha^{3/2}\delta}{\sqrt{\pi}(\delta^2 + a^2)^{3/4}} \exp[\delta\alpha] K_{3/2}[\alpha\sqrt{\delta^2 + a^2}] \quad (10)$$

which, in analogy to (3), can be written as

$$f_A(a) = \int_{\lambda_X} f_{A|\Lambda_X}(a|\lambda_X) f_{\Lambda_X}(\lambda_X) d\lambda_X \quad (11)$$

with $f_{A|\Lambda_X}(a|\lambda_X) = (a/\lambda_X)\exp[-(a^2/2\lambda_X)]$ the Rayleigh distribution and $f_{\Lambda_X}(\lambda_X)$ as in (2).

III. Speech Models and Distributions

In the remaining part of this paper, we consider DFT domain speech enhancement where we assume an additive noise model, i.e., $Y(k,i) = X(k,i) + D(k,i)$, where Y is a noisy speech DFT coefficient, X a clean speech DFT coefficient, D a noise DFT coefficient, k the frequency index, and i the time frame index. The DFT coefficients Y, X, and D are assumed to be complex zero-mean random variables with X and D uncorrelated, e.g., $E[X(k,i)D(k,i)] = 0$. We assume the noise DFT coefficients to be Gaussian distributed, which can be justified by the central limit theorem and the fact that for many noise sources the span of correlation is short compared to the frame size [14].

A. Complex DFT MAP Estimator

The complex DFT MAP estimator that we present is rather general and allows to model speech vector processes and incorporation of correlation between the vector elements. Let $\mathbf{Y} \in \mathbb{R}^d$, $\mathbf{X} \in \mathbb{R}^d$, and $\mathbf{D} \in \mathbb{R}^d$, whose elements can for example be the real or imaginary part of a DFT coefficient X_R or X_I, respectively, at a frequency bin k ($d = 1$), both the real and imaginary DFT coefficients at a frequency bin k ($d = 2$) or even at a series of frequency bins $k = k_1, \ldots, k_2$ ($d = 2(k_2 - k_1 + 1)$).

The MAP estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ is then found by computation of

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) = \arg\max_{\mathbf{x}} \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{y})}. \quad (12)$$

Because $f_{\mathbf{Y}}(\mathbf{y})$ is independent of $\mathbf{x}$ and the natural logarithm is a monotonic function, it is sufficient to maximize

$$\ln f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}). \quad (13)$$

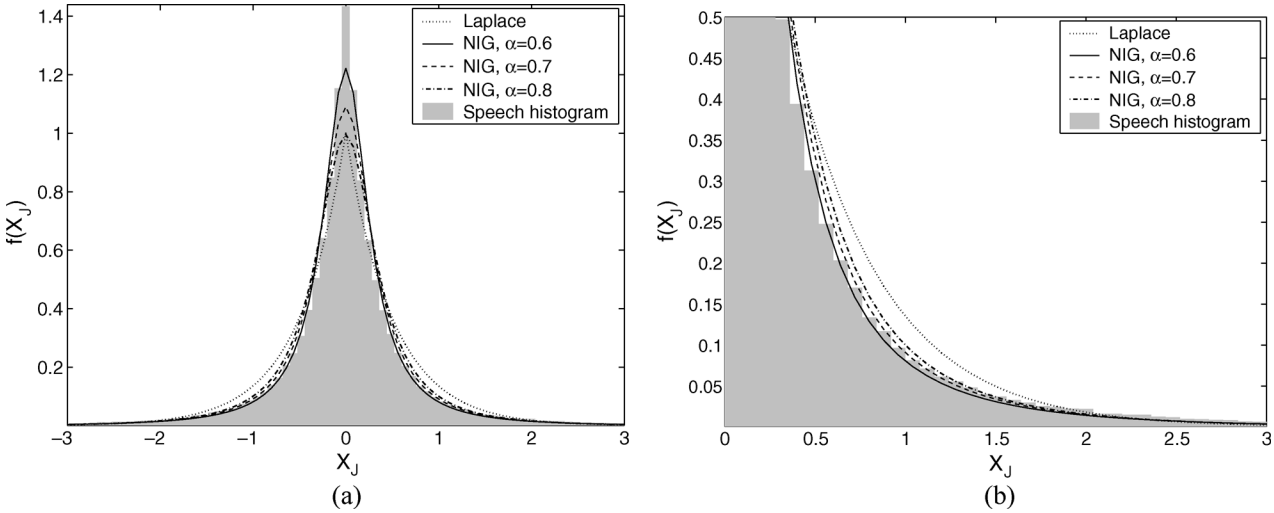


Fig. 3. Histogram of speech DFT coefficients and fitted distributions.

1) A Posteriori Distributions for Complex DFT Coefficients: Under the assumption that $\mathbf{D} \sim N_d(0, \mathbf{\Lambda}_D)$, we can write the distribution of $\mathbf{Y}$ conditioned on $\mathbf{X}$ as

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Lambda}_D|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{y}-\mathbf{x})^T \mathbf{\Lambda}_D^{-1}(\mathbf{y}-\mathbf{x})\right]. \tag{14}$$

We assume that $\mathbf{X}$ is a d-dimensional scale mixture of normals with an MNIG distribution, with $\boldsymbol{\mu} = \boldsymbol{\beta} = 0$, which means that (1) is simplified as $\mathbf{X} = \sqrt{\Lambda_X}\mathbf{\Gamma}^{1/2}\mathbf{Z}$ with distribution

$$f_{\mathbf{X}}(\mathbf{x}) = \int f_{\mathbf{X}|\Lambda_X}(\mathbf{x}|\lambda_X) f_{\Lambda_X}(\lambda_X) d\lambda_X \tag{15}$$

where

$$f_{\mathbf{X}|\Lambda_X}(\mathbf{x}|\lambda_X) = \frac{1}{(2\pi)^{d/2}|\lambda_X\mathbf{\Gamma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2\lambda_X}\mathbf{x}^T\mathbf{\Gamma}^{-1}\mathbf{x}\right] \tag{16}$$

and with the mixing distribution $f_{\Lambda_X}(\lambda_X)$ as in (2), where $\lambda_X\mathbf{\Gamma}$ is the covariance matrix of vector $\mathbf{X}$.

2) Experimental Data: In [2], [3], it was reported that the Laplace distribution provides a much better fit to the histogram of speech DFT coefficients than the Gaussian density. In Fig. 3, we show histograms of the real and imaginary part of speech DFT coefficients that are obtained by selecting only those DFT coefficients that have an estimated *a priori* signal-to-noise ratio (SNR) between 28 and 31 dB. To do so, signal frames of 512 samples where taken with 50% overlap and with a sample frequency of 16 kHz. For all frequency bins, separate histograms were measured, normalized such that $\lambda_X = 1$ and averaged over frequencies. To the histogram, the Laplace and NIG distribution are fitted. Fig. 3 demonstrates that the Laplace and NIG distribution have similar fit around the tails, but that the NIG has a better fit in between the top and the tail of the histogram. Moreover, for the NIG distributions, it is possible to adapt the α parameter to the underlying speech density.

The better fit is also reflected by the estimated Kullback–Leibler discrimination information [15]

$$I_{KB} = \sum_x f_H(x) log\left(\frac{f_H(x)}{f(x)}\right) \tag{17}$$

of the histogram $f_H(x)$ and one of the densities depicted in Fig. 3. It turns out that the Kullback–Leibler discrimination measure is about 4.4 times smaller for an NIG distribution with $\alpha = 0.6$ than for the Laplace distribution.

3) Map Estimator: After substitution of (2) and (14)–(16) in (13) followed by taking the derivative of (13) with respect to $\mathbf{x}$ using rules for matrix calculation and using [16, Th. 3.471,9], we get the derivative f' (see Appendix I)

$$\begin{aligned} f' =& \mathbf{\Lambda}_D^{-1}(\mathbf{y}-\mathbf{x}) \\ & - \mathbf{\Gamma}^{-1}\mathbf{x}\frac{\int_{\lambda_X} \lambda_X^{-1} f_{\mathbf{X}|\Lambda_X}(\mathbf{x}|\lambda_X) f_{\Lambda_X}(\lambda_X) d\lambda_X}{\int_{\lambda_X} f_{\mathbf{X}|\Lambda_X}(\mathbf{x}|\lambda_X) f_{\Lambda_X}(\lambda_X) d\lambda_X} \end{aligned} \tag{18}$$

$$\begin{aligned} =& \mathbf{\Lambda}_D^{-1}(\mathbf{y}-\mathbf{x}) - \mathbf{\Gamma}^{-1}\mathbf{x}\left(\frac{\alpha^2}{\delta^2 + \mathbf{x}^T\mathbf{\Gamma}^{-1}\mathbf{x}}\right)^{\frac{1}{2}} \\ & \times \frac{K_{\frac{3+d}{2}}\left(\sqrt{\alpha^2(\delta^2 + \mathbf{x}^T\mathbf{\Gamma}^{-1}\mathbf{x})}\right)}{K_{\frac{1+d}{2}}\left(\sqrt{\alpha^2(\delta^2 + \mathbf{x}^T\mathbf{\Gamma}^{-1}\mathbf{x})}\right)} \end{aligned} \tag{19}$$

which needs to be solved for $\mathbf{x}$. Unfortunately, it is, by our knowledge, not possible to solve (19) analytically. However, we can use an intermediate step to find an analytic solution. Namely, the ratio of integrals in (18) constitutes an MMSE estimate of the inverse first moment of Λ_X, that is

$$E\left[\Lambda_X^{-1}|\mathbf{x}\right] = \frac{\int_{\lambda_X} \lambda_X^{-1} f_{\mathbf{X}|\Lambda_X}(\mathbf{x}|\lambda_X) f_{\Lambda_X}(\lambda_X) d\lambda_X}{\int_{\lambda_X} f_{\mathbf{X}|\Lambda_X}(\mathbf{x}|\lambda_X) f_{\Lambda_X}(\lambda_X) d\lambda_X}. \tag{20}$$

Assuming that we are given a pre-estimate of $\mathbf{x}$, denoted by $\tilde{\mathbf{x}}$ and using (20) we can solve (18) leading to

$$\hat{\mathbf{x}} = \left(\mathbf{\Lambda}_D^{-1} + E\left[\Lambda_X^{-1}|\tilde{\mathbf{x}}\right]\mathbf{\Gamma}^{-1}\right)^{-1}\mathbf{\Lambda}_D^{-1}\mathbf{y} \tag{21}$$

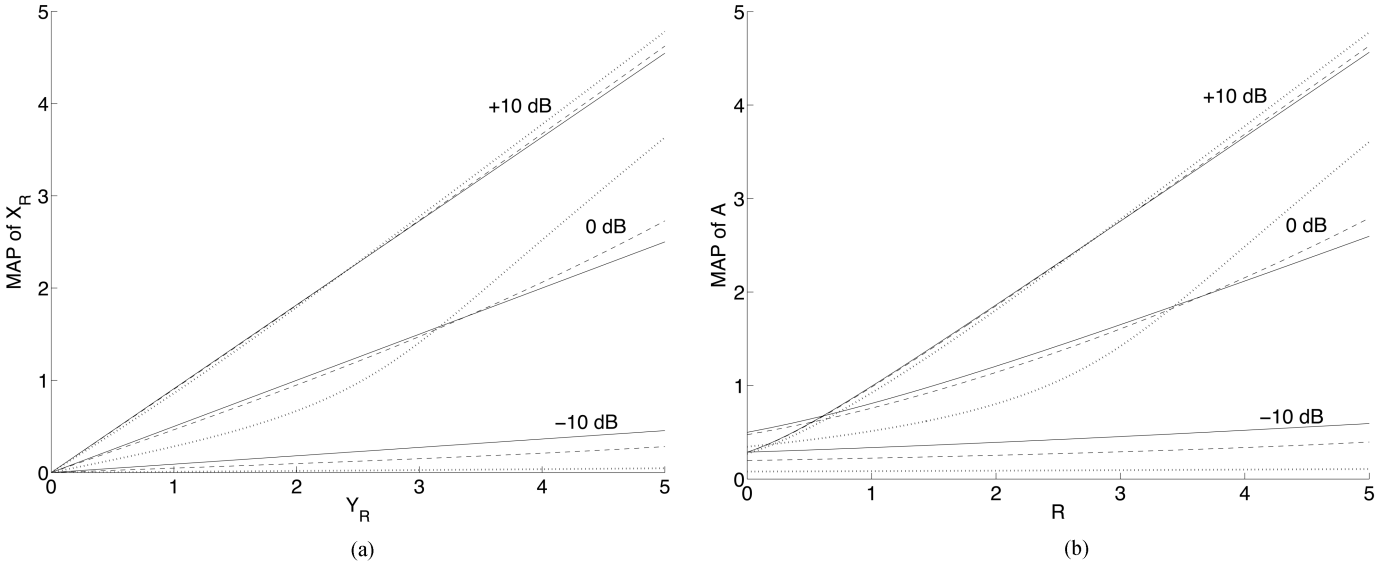


Fig. 4. Input versus output characteristics for *a priori* SNR values 10, 0, and −10 dB with $\Lambda_{X_R} + \Lambda_{D_R} = 2$ and $\alpha = 3$ (dashed) and $\alpha = 1$ (dotted) for (a) the NIG-based estimator compared to the Wiener gain (solid) and (b) the RIG-based estimator compared to the Rayleigh-based amplitude estimator (solid).

with

$$E\left[\Lambda_X^{-1}|\tilde{\mathbf{x}}\right] = \left(\frac{\alpha^2}{\delta^2 + \tilde{\mathbf{x}}^T\boldsymbol{\Gamma}^{-1}\tilde{\mathbf{x}}}\right)^{\frac{1}{2}} \times \frac{K_{\frac{3+d}{2}}\left(\sqrt{\alpha^2(\delta^2 + \tilde{\mathbf{x}}^T\boldsymbol{\Gamma}^{-1}\tilde{\mathbf{x}})}\right)}{K_{\frac{1+d}{2}}\left(\sqrt{\alpha^2(\delta^2 + \tilde{\mathbf{x}}^T\boldsymbol{\Gamma}^{-1}\tilde{\mathbf{x}})}\right)}. \quad (22)$$

For now, we assume $\tilde{\mathbf{x}}$ to be known. In Section IV, we will specify how $\tilde{\mathbf{x}}$ can be obtained in practice. Notice, that when we consider the case $d = 1$ and assume no correlation between real and imaginary parts of DFT coefficients, that $\Lambda_X = VAR[Re(X)] = VAR[Im(X)]$ and $\Lambda_D = VAR[Re(D)] = VAR[Im(D)]$. Further, notice that the Wiener filter is a special case of (21), namely when there is no uncertainty on Λ_X and $f_{\Lambda_X}(\lambda_X)$ becomes a delta impulse.

For $\hat{\mathbf{x}}$ in (21) to constitute a local maximum, it is necessary that the second derivative f'' evaluated at $\hat{\mathbf{x}}$ is negative. Using [16, Th. 8.486,11], the second derivative is given by

$$\begin{aligned} f'' = &-\boldsymbol{\Lambda}_D^{-1} - \frac{K_{\frac{3+d}{2}}(z)}{K_{\frac{1+d}{2}}(z)} \\ &\times \left(\frac{\boldsymbol{\Gamma}^{-1}\alpha}{\sqrt{\delta^2 + \mathbf{x}^T\boldsymbol{\Gamma}^{-1}\mathbf{x}}} - \frac{\alpha\mathbf{x}^T\boldsymbol{\Gamma}^{-2}\mathbf{x}}{(\delta^2 + \mathbf{x}^T\boldsymbol{\Gamma}^{-1}\mathbf{x})^{1\frac{1}{2}}}\right) \\ &- \frac{\mathbf{x}^T\boldsymbol{\Gamma}^{-2}\mathbf{x}\alpha^2}{\delta^2 + \mathbf{x}^T\boldsymbol{\Gamma}^{-1}\mathbf{x}} \\ &\times \frac{K^2_{\frac{3+d}{2}}(z) + K_{\frac{3+d}{2}}(z)K_{\frac{-1+d}{2}}(z) - K^2_{\frac{1+d}{2}}(z)}{2K^2_{\frac{1+d}{2}}(z)} \\ &+ \frac{\mathbf{x}^T\boldsymbol{\Gamma}^{-2}\mathbf{x}\alpha^2}{\delta^2 + \mathbf{x}^T\boldsymbol{\Gamma}^{-1}\mathbf{x}} \frac{K_{\frac{5+d}{2}}(z)}{2K_{\frac{1+d}{2}}(z)} \end{aligned} \quad (23)$$

with $z = \alpha\sqrt{\delta^2 + \mathbf{x}^T\boldsymbol{\Gamma}^{-1}\mathbf{x}}$. Unfortunately, the last term of f'' is positive, which means that for very low noise levels and z close to zero f'' can become positive. For z close to zero both α and $\mathbf{x}^T\boldsymbol{\Gamma}^{-1}\mathbf{x}$ should become close to zero. In practice this will not frequently happen, since a very small α means that the distribution for the clean speech is very heavy-tailed and that large values of $\mathbf{x}^T\boldsymbol{\Gamma}^{-1}\mathbf{x}$ are very likely. In order to overcome this problem, we detect in practice whether f'' is positive. When this occasionally is the case, we assume that there is no uncertainty on Λ_X and replace (2) in (3) with a delta impulse and use (8) for Λ_X, that is $\Lambda_X = (\delta/\alpha)$.

Fig. 4(a) shows the input–output characteristics of (21) when $d = 1$ and applied on the real part of the noisy DFT coefficients Y_R for $0 \leq Y_R \leq 5$ with $\Lambda_X + \Lambda_D = 2$ for the *a priori* SNR values of 10, 0, and −10 dB for various values of α. This is compared with the input–output characteristic of the Wiener filter. Compared to the Wiener filter, the NIG-based estimator shows similar characteristics for an *a priori* SNR of 10 dB and high α values, while for lower α values the NIG-based estimator leads to less suppression for the higher input values. For an *a priori* SNR value of 0 dB, the NIG estimator shows a more pronounced nonlinear characteristics. Compared to the Wiener filter, there is more suppression for smaller input values and less suppression for larger input values. For an *a priori* SNR of −10 dB, the NIG MAP estimator leads to more suppression. Notice that the behavior of the NIG estimator is in principle similar to the super-Gaussian distribution-based estimators as presented in [3], but with the difference that the NIG-based estimator can adapt its shape parameters and as a consequence, its suppression characteristics as well.

B. Amplitude Map Estimator

For the amplitude MAP estimator, we consider 2-D MNIG distributed vectors $\mathbf{X} = [X_R, X_I]^T$ with $\boldsymbol{\Gamma} = \mathbf{I}$. Here, R and I denote the real and imaginary part of a DFT coefficient, respectively, such that $X = X_R + jX_I = Ae^{j\Phi}$ with $j = \sqrt{-1}$. Further, $\mathbf{Y} = [Y_R, Y_I]^T$ such that $Y = Y_R + jY_I = Re^{j\Theta}$

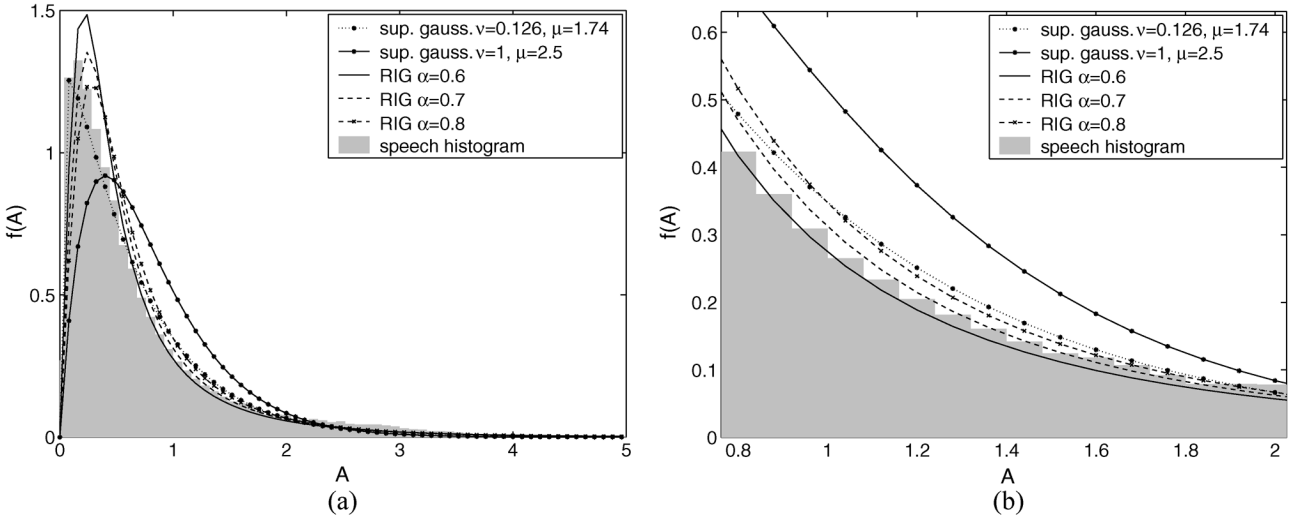


Fig. 5. Histogram of speech DFT amplitudes and fitted distributions.

and $\mathbf{D} = [D_R, D_I]^T$ with $D = D_R + jD_I$. Further, $\Lambda_X = VAR[X_R] = VAR[X_I]$ and $\Lambda_D = VAR[D_R] = VAR[D_I]$.

An amplitude MAP estimator $\hat{a}$ is found by computation of

$$\begin{aligned}\hat{a} &= \arg\max_a f_{A|R}(a|r) \\ &= \arg\max_a \frac{f_{R|A}(r|a) f_A(a)}{f_R(r)}.\end{aligned} \tag{24}$$

Because of the monotonic property of the natural logarithm and the independence of $f_R(r)$ from a, it is sufficient to compute

$$\hat{a} = \arg\max_a \ln\left[f_{R|A}(r|a) f_A(a)\right]. \tag{25}$$

1) A Posteriori Distributions for Amplitudes of Complex DFT Coefficients: The distribution of R given A is given by transformation (14) into polar coordinates, substitution of $Y = R\exp[j\Theta]$ and $X = A\exp[j\Phi]$, and integration over phase

$$f_{R|A}(r|a) = \frac{r}{\Lambda_D}\exp\left[-\frac{r^2+a^2}{2\Lambda_D}\right] I_0\left(\frac{ar}{\Lambda_D}\right). \tag{26}$$

For $(ar/\Lambda_D) \geq 3$, it is reasonable to approximate $I_0(x)$ by $I_0 \approx (1/\sqrt{2\pi x})\exp[x]$ [17], such that

$$f_{R|A}(r|a) = \frac{r}{\Lambda_D}\exp\left[-\frac{r^2-2ar+a^2}{2\Lambda_D}\right]\sqrt{\frac{\Lambda_D}{2\pi ar}}. \tag{27}$$

We assume that the speech amplitudes A are RIG distributed and its distribution is given by (10) and (11).

2) Experimental Data: In Fig. 5, a histogram measured over amplitudes is depicted. The data for this histogram is obtained similarly as for Fig. 3. To this histogram, the RIG distribution for several α values and the super-Gaussian approximations defined in [4] as

$$f_A(a) = \frac{\mu^{\nu+1}a^\nu}{\Gamma(\nu+1)(2\lambda_X)^{\frac{\nu+1}{2}}}\exp\left[-\mu\frac{a}{\sqrt{2\lambda_X}}\right] \tag{28}$$

are fitted, with parameters $(\nu,\mu) = (1, 2.5)$ and $(\nu,\mu) = (0.126, 1.74)$ as given in [4], where the latter set was chosen in [4] to optimize for the used dataset in [4]. Especially for amplitudes in the range of $1 \leq A \leq 2$, the RIG distribution shows a better fit than the two super-Gaussian approximations. Moreover, due to its flexibility, the RIG can be adapted to the underlying speech distribution.

The Kullback–Leibler discrimination measure (17) is about the same for the RIG distribution with $\alpha = 0.6$ and the super-Gaussian distribution with parameter settings $(\nu,\mu) = (0.126, 1.74)$. Compared to the super-Gaussian distribution with parameter settings $(\nu,\mu) = (1, 2.5)$, the RIG distribution with $\alpha = 0.6$ has a Kullback–Leibler discrimination measure that is more than seven times smaller.

3) MAP Estimator for Amplitudes: Substitution of (10) and (27) in (25) and taking the derivative with respect to a using [16, Th. 3.471,9] gives the derivative f' (see Appendix II)

$$f' = \frac{-a+r}{\Lambda_D} + \frac{1}{2a} - a\frac{\int \lambda_X^{-1} f_{A|\Lambda_X}(a|\lambda_X) f_{\Lambda_X}(\lambda_X) d\lambda_X}{\int f_{A|\Lambda_X}(a|\lambda_X) f_{\Lambda_X}(\lambda_X) d\lambda_X} \tag{29}$$

$$= \frac{-a+r}{\Lambda_D} + \frac{1}{2a} - a\left(\frac{\alpha^2}{\delta^2+a^2}\right)^{\frac{1}{2}}\frac{K_{2\frac{1}{2}}\left(\alpha\sqrt{\delta^2+a^2}\right)}{K_{1\frac{1}{2}}\left(\alpha\sqrt{\delta^2+a^2}\right)}. \tag{30}$$

The amplitude MAP estimator is then given by solving (30) for a. Similar as for the complex DFT MAP estimator, it is not possible to solve this equation analytically for a. Therefore, we use the intermediate solution where the ratio of integrals in (29) constitutes an MMSE estimate of the inverse first moment of Λ_X, that is

$$E\left[\Lambda_X^{-1}|a\right] = \left(\frac{\alpha^2}{\delta^2+a^2}\right)^{\frac{1}{2}}\frac{K_{2\frac{1}{2}}\left(\alpha\sqrt{\delta^2+a^2}\right)}{K_{1\frac{1}{2}}\left(\alpha\sqrt{\delta^2+a^2}\right)}. \tag{31}$$

Given a pre-estimate of a, denoted by $\tilde{a}$ and using (31), we solve (29) leading to

$$\hat{a} = \frac{\frac{1}{\Lambda_D} + \sqrt{\frac{1}{\Lambda_D^2} + 2\left(\frac{1}{r^2\Lambda_D} + \frac{E\left[\Lambda_X^{-1}|\tilde{a}\right]}{r^2}\right)}}{2\left(\frac{1}{\Lambda_D} + E\left[\Lambda_X^{-1}|\tilde{a}\right]\right)} r \tag{32}$$

where the second solution of (29) is neglected, because that leads to $a < 0$.

Notice, that the MAP amplitude estimator proposed in [18] is a special case of (32), namely, when there is no uncertainty on λ_X, and $f_{\Lambda_X}(\lambda_X)$ becomes a delta impulse.

The second derivative of the amplitude estimator is using [16, Th. 8.486,11], given by (33) at the bottom of the page, with $z = \alpha\sqrt{\delta^2 + a^2}$. Although unlikely to happen, in practice, f'' can become positive when z is very small and the noise level very high. When occasionally this is the case, we take the same measures as mentioned in Section III-A3.

In Fig. 4(b), the input–output characteristics of the RIG amplitude estimator are shown and compared with the characteristics of the MAP estimator under the Rayleigh distribution as presented in [18]. The characteristics are normalized by $\Lambda_X + \Lambda_D = 2$. The exact characteristics of the RIG-based estimator depend on the α parameter. When α gets larger, the characteristics are close to the Rayleigh distribution-based estimator, while for smaller α values, the characteristics show less suppression for the larger input values, which will preserve speech components, and more suppression to the lower input values which leads to more noise reduction.

IV. Experimental Results

The performance of the presented clean speech estimators is evaluated in terms of segmental SNR defined as [19]

$$\mathrm{SNR}_{\mathrm{seg}} = \frac{1}{\mathcal{L}} \sum_{i=0}^{\mathcal{L}-1} \mathcal{T}\left\{ 10\log_{10} \frac{\|\mathbf{x}_t(i)\|^2}{\|\mathbf{x}_t(i) - \hat{\mathbf{x}}_t(i)\|^2} \right\} \tag{34}$$

where $\mathbf{x}_t(i)$ and $\hat{\mathbf{x}}_\mathbf{t}(i)$ are time-domain vectors and denote frame i of the clean speech signal $\mathbf{x}_t$ and the enhanced speech signal $\hat{\mathbf{x}}_t$, respectively. $\mathcal{L}$ is the number of frames within the speech signal in question and $\mathcal{T}(z) = \min\{\max(z, -10), 35\}$ a function which limits the SNR to a perceptually meaningful range.

In addition to $\mathrm{SNR}_{\mathrm{seg}}$, we use intelligibility weighted segmental SNR [20], defined as

$$\mathrm{IWSNR}_{\mathrm{seg}} = \frac{1}{\mathcal{L}} \sum_{i=0}^{\mathcal{L}-1} \sum_{b=1}^{B} w(b)\mathcal{T} \times \left\{ 10\log_{10} \frac{\|\mathbf{x}_t(i,b)\|^2}{\|\mathbf{x}_t(i,b) - \hat{\mathbf{x}}_t(i,b)\|^2} \right\} \tag{35}$$

where the weight $w(b)$ emphasizes the importance of the bth frequency band, and where $\mathbf{x}_t(i,b)$ is a time domain vector in band b.

To get an indication whether a difference in $\mathrm{SNR}_{\mathrm{seg}}$ is due to more noise reduction or less speech distortion, we process the clean signal and the noise signal with the same filters $\mathbf{G}$ as used to enhance the noisy signal and measure noise attenuation $\mathrm{NATT}_{\mathrm{seg}}$, defined as

$$\mathrm{NATT}_{\mathrm{seg}} = \frac{1}{\mathcal{L}} \sum_{i=0}^{\mathcal{L}-1} 10\log_{10} \frac{\|\mathbf{n}_t(i)\|^2}{\|\mathbf{G}(i)\mathbf{n}_t(i)\|^2} \tag{36}$$

where $\mathbf{n}(i)$ is a vector and denotes frame i of the noise sequence, and $\mathbf{G}(i)$ is a filtering matrix. Speech attenuation is defined as

$$\mathrm{SATT}_{\mathrm{seg}} = \frac{1}{\mathcal{L}} \sum_{i=0}^{\mathcal{L}-1} 10\log_{10} \frac{\|\mathbf{x}_t(i)\|^2}{\|\mathbf{x}_t(i) - \mathbf{G}(i)\mathbf{x}_t(i)\|^2}. \tag{37}$$

For both $\mathrm{NATT}_{\mathrm{seg}}$ and $\mathrm{SATT}_{\mathrm{seg}}$, only those frame are taken into account where the SNR of the noisy frame i is larger than -10 dB.

Experiments are done with speech signals degraded with white noise, F16 noise, car noise, and Factory noise, at the input SNR ratios of 5, 10, and 15 dB. The speech and noise signals originate from the TIMIT database [21] and NOISEX 92 database [22], respectively. All results are averaged over 32 different speech signals all sampled at 16 kHz. We use a frame size of 512 samples with 50% overlap of adjacent frames. Noise statistics are measured using the minimum statistics approach [23]. For all methods that we compare in this section, we set a lower bound to the enhancement gain of -15 dB. To compute (22) and (31), we make use of a preliminary estimate of the clean signal by applying a Wiener filter to the noisy speech signal. The parameters α and δ in (22) and (31) are computed per frame and per frequency bin using the expectation-maximization procedure presented in [11].

To evaluate the performance of the MNIG MAP estimator, we compare (21) with $d = 1$ with the Laplace-based MMSE estimator [3] and the Wiener filter. The results in Table I show that the improvement of the NIG MAP estimator compared to the Laplace-based MMSE estimator varies, dependent on noise source and noise level, from 0.3 to 0.6 dB. Compared to the Wiener filter, the improvement varies from 0.6 to 1.4 dB.

The performance of the RIG-based amplitude MAP estimator in (32) is compared with the Rayleigh distribution-based MAP amplitude estimator proposed in [18] and the super-Gaussian MAP amplitude estimator and the joint MAP amplitude and phase estimator as proposed in [4], with the distribution as in (28) with $(\nu, \mu) = (1, 2.5)$ and $(\nu, \mu) = (0.126, 1.74)$, abbreviated with supergauss^1 and supergauss^2, respectively. Table I shows that the RIG-based MAP estimator has an improvement in terms of $\mathrm{SNR}_{\mathrm{seg}}$ of 0.2 to 0.6 dB compared to the supergauss^1 estimator and an improvement in terms of $\mathrm{SNR}_{\mathrm{seg}}$ of 0.2 to 0.4 dB compared to the supergauss^2 estimator. Compared to the estimator under the Rayleigh distribution, the improvement in terms of $\mathrm{SNR}_{\mathrm{seg}}$ is in the order of 0.8 to 1.6 dB.

$$f'' = -\frac{1}{\Lambda_D} - \frac{1}{2a^2} - \frac{K_2\frac{1}{2}(\alpha\sqrt{\delta^2+a^2})}{K_1\frac{1}{2}(\alpha\sqrt{\delta^2+a^2})} \left(\frac{\alpha}{\sqrt{\delta^2+a^2}} - \frac{\alpha a^2}{\sqrt{\delta^2+a^2}(\delta^2+a^2)} \right) - \frac{\alpha^2 a^2}{\delta^2+a^2} \frac{K_2\frac{1}{2}(z)^2 + K_2\frac{1}{2}(z)K_{\frac{1}{2}}(z) - K_1\frac{1}{2}(z)^2 - K_1\frac{1}{2}(z)K_3\frac{1}{2}(z)}{2K_1\frac{1}{2}(z)^2} \tag{33}$$

TABLE I
IMPROVEMENT IN $\mathrm{SNR}_{\mathrm{seg}}$ (dB)

input noise source	input SNR (dB)	$\mathrm{SNR}_{\mathrm{seg}}$ (dB) DFT est. NIG	Wiener	Laplace	$\mathrm{SNR}_{\mathrm{seg}}$ (dB) Amplitude est. RIG	Rayleigh	Super-gauss[1]	Super-gauss[2]
white	5	6.2	5.2	5.7	6.1	4.9	5.7	5.9
	10	5.3	4.1	4.8	5.2	3.8	4.7	4.9
	15	4.3	2.9	3.7	4.3	2.7	3.7	3.9
F16	5	5.3	4.3	4.8	5.2	4.0	4.8	5.0
	10	4.5	3.4	4.0	4.4	3.1	4.0	4.1
	15	3.7	2.4	3.2	3.6	2.2	3.2	3.3
Car	5	7.0	6.1	6.5	6.9	5.6	6.6	6.6
	10	5.5	4.5	5.0	5.4	4.2	5.1	5.1
	15	3.7	2.6	3.2	3.7	2.5	3.2	3.3
Factory	5	3.8	3.2	3.6	3.8	2.9	3.5	3.6
	10	3.4	2.4	3.0	3.3	2.2	2.9	3.0
	15	2.9	1.7	2.4	2.8	1.5	2.3	2.5

TABLE II
IMPROVEMENT IN $\mathrm{IWSNR}_{\mathrm{seg}}$ (dB)

input noise source	input SNR (dB)	$\mathrm{IWSNR}_{\mathrm{seg}}$ (dB) DFT est. NIG	Wiener	Laplace	$\mathrm{IWSNR}_{\mathrm{seg}}$ (dB) Amplitude est. RIG	Rayleigh	Super-gauss[1]	Super-gauss[2]
white	5	4.8	4.1	4.4	4.7	3.4	4.5	4.6
	10	4.3	3.5	3.9	4.2	2.9	3.9	4.0
	15	3.5	2.6	3.1	3.5	2.1	3.1	3.2
F16	5	4.6	3.9	4.2	4.5	3.2	4.3	4.4
	10	4.1	3.1	3.6	3.9	2.6	3.7	3.8
	15	3.3	2.3	2.9	3.3	1.9	2.9	3.0
Car	5	-0.58	-1.8	-1.1	-0.51	-1.7	-1.1	-0.90
	10	-1.8	-3.0	-2.3	-1.6	-2.8	-2.3	-2.1
	15	-3.1	-4.3	-3.6	-2.8	-3.9	-3.6	-3.4
Factory	5	3.6	3.0	3.3	3.5	2.4	3.4	3.3
	10	3.2	2.4	2.8	3.1	1.9	2.8	2.9
	15	2.7	1.6	2.2	2.6	1.3	2.2	2.3

TABLE III
$\mathrm{SATT}_{\mathrm{seg}}$ (dB)

input noise source	input SNR (dB)	$\mathrm{SATT}_{\mathrm{seg}}$ (dB) DFT est. NIG	Wiener	Laplace	$\mathrm{SATT}_{\mathrm{seg}}$ (dB) Amplitude est. RIG	Rayleigh	Supergauss[1]	Supergauss[2]
white	5	10.0	8.2	9.3	10.3	9.0	9.1	9.7
	10	12.7	10.6	11.8	13.1	11.5	11.7	12.4
	15	15.7	13.2	14.6	16.1	14.1	14.5	15.2
F16	5	10.1	8.0	9.2	10.4	9.0	9.1	9.8
	10	13.4	10.8	12.2	13.7	11.9	12.1	12.9
	15	16.9	13.8	15.4	17.3	15.0	15.3	16.2
Car	5	23.8	20.7	21.9	24.2	22.2	22.0	22.8
	10	26.7	23.1	24.4	27.1	24.6	24.5	25.3
	15	28.9	25.2	26.5	29.4	26.5	26.6	27.4
Factory	5	10.4	8.2	9.5	10.7	9.4	9.3	10.1
	10	13.8	11.0	12.5	14.2	12.2	12.4	13.3
	15	17.6	14.1	15.7	17.9	15.3	15.7	16.6

In addition to segmental SNR, we show in Table II the performance improvement in terms of intelligibility weighted segmental SNR. The MNIG MAP estimator has an improvement in terms of $\mathrm{IWSNR}_{\mathrm{seg}}$ of 0.3 to 0.5 dB compared to the Laplace-based estimator and 0.6 to 1.2 dB compared to the Wiener filter. The RIG based amplitude MAP estimator has an improvement in terms of $\mathrm{IWSNR}_{\mathrm{seg}}$ of and 0.1 to 0.7 dB and 0.1 to 0.5 dB compared to supergauss[1] and supergauss[2], respectively, and an improvement of 1.1 to 1.3 dB compared to the estimator under the Rayleigh distribution.

In Tables III and IV, we show the $\mathrm{SATT}_{\mathrm{seg}}$ and $\mathrm{NATT}_{\mathrm{seg}}$ scores, respectively. It reveals that the proposed NIG and RIG MAP estimators in general have a better speech quality in terms of $\mathrm{SATT}_{\mathrm{seg}}$ due to the adaptation of the distribution per frequency bin, but a somewhat less noise reduction than the estimators based on preselected super-Gaussian densities.

Informal listening revealed that the proposed MNIG filter has a tendency to produce slightly more musical tones than the Wiener filter, but that speech sounds are better preserved and sound less suppressed.

TABLE IV
$\mathrm{NATT}_{\mathrm{seg}}$ (dB)

input noise source	SNR (dB)	$\mathrm{NATT}_{\mathrm{seg}}$ (dB) DFT est. NIG	Wiener	Laplace	$\mathrm{NATT}_{\mathrm{seg}}$ (dB) Amplitude est. RIG	Rayleigh	Supergauss[1]	Supergauss[2]
white	5	14.7	16.1	16.3	13.6	12.9	15.1	14.3
	10	11.9	13.3	13.5	10.9	10.4	12.2	11.6
	15	9.7	10.8	11.3	8.7	8.2	9.8	9.4
F16	5	12.3	13.8	14.0	11.2	10.7	12.8	12.0
	10	9.8	11.2	11.6	8.9	8.5	10.2	9.7
	15	8.0	9.1	9.7	7.0	6.8	8.3	7.9
Car	5	12.5	13.6	13.7	11.6	10.8	12.9	12.3
	10	11.0	12.1	12.4	10.1	9.5	11.4	10.9
	15	9.9	10.8	11.3	8.8	8.2	10.2	9.8
Factory	5	9.7	11.5	11.2	8.9	8.7	10.3	9.5
	10	8.0	9.6	9.7	7.2	7.1	8.5	8.0
	15	6.7	8.0	8.4	5.9	5.8	7.1	6.7

Notice that estimation of δ and α for the presented algorithms with the EM algorithm as presented in [11] is computationally demanding. However, in more recent research, we developed a more simplified way to estimate δ and α without the use of the EM algorithm. Preliminary experiments based on the same data-set degraded with white noise at SNRs of 5, 10, and 15 dB showed similar performance as the results presented in this paper.

V. Conclusion

In this paper, we presented a new class of complex DFT and amplitude estimators for DFT domain-based speech enhancement. The estimators are derived under a multidimensional normal inverse Gaussian (MNIG) distribution for the DFT coefficients. As a scale mixture, the MNIG distribution is very flexible and can model a wide range of densities, from heavy-tailed to less heavy-tailed. Under the MNIG distribution, we derived complex DFT and amplitude estimators, where the suppression characteristics of the estimators can be adapted online to the distribution of the observed speech DFT coefficients. Measurements of speech histograms based on speech DFT coefficients and DFT amplitudes showed a slightly better fit for the MNIG and RIG distribution, respectively, than for the preselected super-Gaussian distributions. Experimental results demonstrated improvement in comparison to complex DFT and amplitude estimators that are based on Gaussian and preselected super-Gaussian distributions. Further, the derived complex MNIG-based estimator allows for vector processing, where correlation between vector elements can be taken into account.

Appendix I

Here, we outline the steps to compute the MAP estimator under the NIG distribution for the complex DFT coefficients. The derivative of $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})$ can be computed as

$$\begin{aligned}
&\frac{d}{d\mathbf{x}^T}\ln\left[f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})\right]\\
&=\frac{d}{d\mathbf{x}^T}\ln\left[f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})\right]+\frac{d}{d\mathbf{x}^T}\ln\left[f_{\mathbf{X}}(\mathbf{x})\right]\\
&=\boldsymbol{\Lambda}_{\mathbf{D}}^{-1}(\mathbf{y}-\mathbf{x})+\frac{\frac{d}{d\mathbf{x}^T}\int_{\lambda_X} f_{\mathbf{X}|\Lambda_X}(\mathbf{x}|\lambda_X)f_{\Lambda_X}(\lambda_X)d\lambda_X}{\int_{\lambda_X} f_{\mathbf{X}|\Lambda_X}(\mathbf{x}|\lambda_X)f_{\Lambda_X}(\lambda_X)d\lambda_X}\\
&=\boldsymbol{\Lambda}_{\mathbf{D}}^{-1}(\mathbf{y}-\mathbf{x})-\boldsymbol{\Gamma}^{-1}\mathbf{x}\frac{\int_{\lambda_X}\lambda_X^{-1} f_{\mathbf{X}|\Lambda_X}(\mathbf{x}|\lambda_X)f_{\Lambda_X}(\lambda_X)d\lambda_X}{\int_{\lambda_X} f_{\mathbf{X}|\Lambda_X}(\mathbf{x}|\lambda_X)f_{\Lambda_X}(\lambda_X)d\lambda_X}\\
&=\boldsymbol{\Lambda}_{\mathbf{D}}^{-1}(\mathbf{y}-\mathbf{x})-E\left[\Lambda_X^{-1}|\mathbf{x}\right]\boldsymbol{\Gamma}^{-1}\mathbf{x}. \qquad (38)
\end{aligned}$$

Solving

$$\frac{d}{d\mathbf{x}^T}\ln\left[f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})\right]=0$$

then leads to (21). Further, using [16, Th. 3.471,9], we show in (39) at the bottom of the previous page how $E\left[\Lambda_X^{-1}\mathbf{x}\right]$ can be computed, where $K_{d'}$ denotes the modified Bessel function of the second kind with order d'.

$$\begin{aligned}
E\left[\Lambda_X^{-1}|\mathbf{x}\right]&=\frac{\int_{\lambda_X}\lambda_X^{-1} f_{\mathbf{X}|\Lambda_X}(\mathbf{x}|\lambda_X)f_{\Lambda_X}(\lambda_X)d\lambda_X}{\int_{\lambda_X} f_{\mathbf{X}|\Lambda_X}(\mathbf{x}|\lambda_X)f_{\Lambda_X}(\lambda_X)d\lambda_X}\\
&=\frac{\int_{\lambda_X=0}^{\infty}\lambda_X^{-2\frac{1}{2}-d/2}\exp\left[-\frac{1}{2}\left((\delta^2+\mathbf{x}^T\boldsymbol{\Gamma}^{-1}\mathbf{x})\lambda_X^{-1}+\alpha^2\lambda_X\right)\right]d\lambda_X}{\int_{\lambda_X=0}^{\infty}\lambda_X^{-1\frac{1}{2}-d/2}\exp\left[-\frac{1}{2}\left((\delta^2+\mathbf{x}^T\boldsymbol{\Gamma}^{-1}\mathbf{x})\lambda_X^{-1}+\alpha^2\lambda_X\right)\right]d\lambda_X}\\
&=\left(\frac{\alpha^2}{\delta^2+\mathbf{x}^T\boldsymbol{\Gamma}^{-1}\mathbf{x}}\right)^{\frac{1}{2}}\frac{K_{\frac{3+d}{2}}\left(\sqrt{\alpha^2(\delta^2+\mathbf{x}^T\boldsymbol{\Gamma}^{-1}\mathbf{x})}\right)}{K_{\frac{1+d}{2}}\left(\sqrt{\alpha^2(\delta^2+\mathbf{x}^T\boldsymbol{\Gamma}^{-1}\mathbf{x})}\right)}, \qquad (39)
\end{aligned}$$

APPENDIX II

Here, we outline the steps to compute the MAP estimator under the RIG distribution for the DFT amplitudes coefficients. The derivative of $f_{A|R}(a|r)$ can be computed as

$$\begin{aligned}
&\frac{d}{da}\ln\left[f_{A|R}(a|r)\right] \\
&\quad= \frac{d}{da}\ln\left[f_{R|A}(r|a)\right] + \frac{d}{da}\ln\left[f_A(a)\right] \\
&\quad= \frac{-a+r}{\Lambda_D} - \frac{1}{2a} + \frac{\int \frac{d}{da} f_{A|\Lambda_X}(a|\lambda_X) f_{\Lambda_X}(\lambda_X) d\lambda_X}{\int f_{A|\Lambda_X}(a|\lambda_X) f_{\Lambda_X}(\lambda_X) d\lambda_X} \\
&\quad= \frac{-a+r}{\Lambda_D} + \frac{1}{2a} - a\frac{\int \lambda_X^{-1} f_{A|\Lambda_X}(a|\lambda_X) f_{\Lambda_X}(\lambda_X) d\lambda_X}{\int f_{A|\Lambda_X}(a|\lambda_X) f_{\Lambda_X}(\lambda_X) d\lambda_X} \\
&\quad= \frac{-a+r}{\Lambda_D} + \frac{1}{2a} - aE\left[\Lambda_X^{-1}|a\right]. \qquad (40)
\end{aligned}$$

Solving

$$\frac{d}{da}\ln\left[f_{A|R}(a|r)\right] = 0 \qquad (41)$$

then leads to (32). Further, using [16, Th. 3.471,9], it can be shown that

$$\begin{aligned}
&E\left[\Lambda_X^{-1}|a\right] \\
&\quad= \frac{\int \lambda_X^{-1} f_{A|\Lambda_X}(a|\lambda_X) f_{\Lambda}(\lambda_X) d\lambda_X}{\int f_{A|\Lambda_X}(a|\lambda_X) f_{\Lambda_X}(\lambda_X) d\lambda_X} \\
&\quad= \frac{\int \lambda_X^{-3\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\delta^2+a^2\right)\lambda_X^{-1} - \frac{1}{2}\alpha^2\lambda_X\right] d\lambda_X}{\int \lambda_X^{-2\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\delta^2+a^2\right)\lambda_X^{-1} - \frac{1}{2}\alpha^2\lambda_X\right] d\lambda_X} \\
&\quad= \left(\frac{\alpha^2}{\delta^2+a^2}\right)^{\frac{1}{2}} \frac{K_{2\frac{1}{2}}\left(\alpha\sqrt{\delta^2+a^2}\right)}{K_{1\frac{1}{2}}\left(\alpha\sqrt{\delta^2+a^2}\right)} \qquad (42)
\end{aligned}$$

where $K_{d'}$ denotes the modified Bessel function of the second kind with order d'.

REFERENCES

[1] R. Martin, "Speech enhancement using MMSE short time spectral estimation with gamma distributed speech priors," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2002, vol. 1, pp. 253–256.

[2] R. Martin and C. Breithaupt, "Speech enhancement in the DFT domain using Laplacian speech priors," in *Proc. Int. Workshop Acoust., Echo, Noise Control*, Sep. 2003, pp. 87–90.

[3] R. Martin, "Speech enhancement based on minimum mean-square error estimation and super-Gaussian priors," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 845–856, Sep. 2005.

[4] T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-gaussian speech model," *EURASIP J. Appl. Signal Process.*, vol. 7, pp. 1110–1126, May 2005.

[5] J. Jensen, I. Batina, R. C. Hendriks, and R. Heusdens, "A study of the distribution of time-domain speech samples and discrete fourier coefficients," in *Proc. IEEE First BENELUX/DSP Valley Signal Process. Symp.*, Apr. 2005, pp. 155–158.

[6] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Process. Lett.*, vol. 10, no. 7, pp. 204–207, Jul. 2003.

[7] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

[8] D. F. Andrews and C. L. Mallows, "Scale mixtures of normal distributions," *J. R. Statist. Soc.*, ser. B, vol. 36, no. 1, pp. 99–102, 1974.

[9] M. West, "On scale mixtures of normal distributions," *Biometrika*, vol. 74, no. 3, pp. 646–648, 1987.

[10] O. E. Barndorff-Nielsen, "Normal inverse Gaussian distributions and stochastic volatility modelling," *Scand. J. Statist.*, vol. 24, pp. 1–13, 1997.

[11] T. A. Øigård, A. Hanssen, R. E. Hansen, and F. Godtliebsen, "EM-estimation and modeling of heavy-tailed processes with the multivariate normal inverse Gaussian distribution," *Signal Process.*, vol. 85, pp. 1655–1673, 2005.

[12] M. Tweedie, "Functions of a statistical variate with given means with special reference to laplacian distributions," *Proc. Cambridge Philos. Soc.*, vol. 43, pp. 41–49, 1947.

[13] T. Eltoft, "The Rician inverse Gaussian distribution: a new model for non-Rayleigh signal amplitude statistics," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1722–1735, Nov. 2005.

[14] D. Brillinger, *Time Series: Data Analysis and Theory*. San Francisco, CA: Holden-Day, 1981.

[15] S. Kullback, *Information Therory and Statistics*. New York: Dover, 1997.

[16] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products*, 6th ed. New York: Academic, 2000.

[17] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, no. 2, pp. 137–145, Apr. 1980.

[18] P. J. Wolfe and S. J. Godsill, "Efficient alternatives to the Ephraim and Malah suppressionrule for audio signal enahncement," *EURASIP J. Appl. Signal Process.*, vol. 10, pp. 1043–1051, 2003.

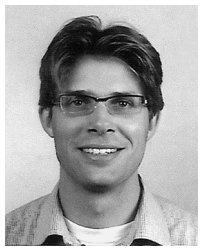
[19] J. Deller, J. H. L. Hansen, and J. G. Proakis, *Discrete-Time Processing of Speech Signals*. Piscataway, NJ: IEEE Press, 2000.

[20] J. Greenberg, P. Peterson, and P. Zurek, "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance," *J. Acoust. Soc. Amer.*, vol. 94, pp. 3009–3010, 1993.

[21] *TIMIT, Acoustic-Phonetic Continuous Speech Corpus*, NIST Speech Disc 1-1.1, DARPA, Oct. 1990.

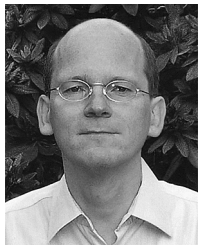
[22] A. Varga and H. J. M. Steeneken, *Noisex-92: A Database and an Experiment to Study the Effect of Additive Noise on Speech Recognition Systems*, vol. 12, no. 3, pp. 247–253, 1993.

[23] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, Jul. 2001.



Richard C. Hendriks received the B.Sc. and M.Sc. degrees, both in electrical engineering, from Delft University of Technology, Delft, The Netherlands, in 2001 and 2003, respectively. He is currently pursuing the Ph.D. degree in the Department of Mediamatics, Delft University of Technology.

His main interests are digital speech and audio processing, including acoustical noise reduction and speech enhancement.



Rainer Martin (S'86–M'90–SM'01) received the Dipl.-Ing. and Dr.-Ing. degrees from Aachen University of Technology, Aachen, Germany, in 1988 and 1996, respectively, and the M.S.E.E. degree from Georgia Institute of Technology, Atlanta, in 1989.

From 1996 to 2002, he has been a Senior Research Engineer with the Institute of Communication Systems and Data Processing, Aachen University of Technology. From April 1998 to March 1999, he was on leave to the AT&T Speech and Image Processing Services Research Laboratory, Florham Park, NJ. From April 2002 until October 2003, he was a Professor of Digital Signal Processing at the Technical University of Braunschweig, Braunschweig, Germany. Since October 2003, he has been a Professor of Information Technology at Ruhr-University Bochum, Bochum, Germany, and Head of the Institute of Communication Acoustics, Bochum. His research interests are signal processing for voice communication systems, acoustics, and human–machine interfaces.