Symmetric $\alpha-$Stable Sparse Linear Regression for Musical Audio Denoising

Nikoletta Bassiou
Department of Informatics
Aristotle University of Thessaloniki
Box 451, Thessaloniki 541 24
GREECE
Email: nbassiou@aiia.csd.auth.gr

Constantine Kotropoulos
Department of Informatics
Aristotle University of Thessaloniki
Box 451, Thessaloniki 541 24
GREECE
Email: costas@aiia.csd.auth.gr

Evangelia Koliopoulou
Department of Informatics
Aristotle University of Thessaloniki
Box 451, Thessaloniki 541 24
GREECE
Email: eyaklp@gmail.com

***Abstract*—A new musical audio denoising technique is proposed, when the noise is modeled by an $\alpha-$stable distribution. The proposed technique is based on sparse linear regression with structured priors and uses Markov Chain Monte Carlo inference to estimate the clean signal model parameters and the $\alpha-$stable noise model parameters. Experiments on noisy Greek folk music excerpts demonstrate better denoising for the $\alpha-$stable noise assumption than the Gaussian white noise one.**

I. Introduction

Over the past decades, signal processing applications have been widely based on the Gaussian assumption. However, there are applications, where this assumption does not hold. Such applications entail non-Gaussian phenomena that exhibit outliers, impulsiveness, and asymmetric characteristics [1], [2], [3], [4]. Recently, $\alpha-$stable distributions have been employed to model these phenomena. The $\alpha-$stable distributions possess several useful properties, including infinite variance, skewness, and heavy tails [5], [6], [7]. However, research has been mostly focused on symmetric $\alpha-$stable distributions within a Bayesian framework, since the probability density function (PDF) for $\alpha-$stable distributions cannot be analytically described in general. In [8], a particular mathematical representation was exploited to infer the $\alpha-$stable parameters using the Gibbs sampler, while in [9], [10] Monte Carlo Expectation-Maximization (MCEM) and Markov Chain Monte Carlo (MCMC) methods were introduced, which are based on the Scale Mixture of Normals (SMiN) representation of $\alpha-$stable distributions. The SMiN property was also exploited to model symmetric $\alpha-$stable (SaS) disturbances by a Gibbs Metropolis sampler [11]. More recently, a random walk MCMC approach for Bayesian inference in stable distributions was introduced using a numerical approximation of the likelihood function [12]. An analytical approximation of positive $\alpha-$stable distribution based on a decomposition into a product of a Pearson and another positive stable random variable was proposed in [13]. Finally, a Poisson sum series representation for the symmetric (SaS) distribution was used to express a noise process in a conditionally Gaussian framework [4].

In this paper, the noise in musical audio recordings is modeled by an $\alpha-$stable distribution. MCMC inference is used to estimate the signal and the $\alpha-$stable noise parameters following similar lines to [12], [14]. The signal is modeled in the frequency domain using the modified cosine transform (MDCT) [15] allowing us to exploit the sparsity in the coefficient expansion. That is, the signal is modeled by two MDCT bases. The first MDCT base describes the tonal parts of the signal, while the second one describes its transient parts as in [16]. However, here the residual noise is treated as $\alpha-$stable noise extending [16] where a Gaussian white noise was assumed. Binary indicator variables with structured priors are also introduced to enforce sparsity in the expansion coefficients of each MDCT base. A standard MCMC technique is used to infer the model parameters. The first experimental results, reported here, demonstrate a superior performance for the $\alpha-$stable noise with respect to the power of the remaining noise after denoising and the acoustic perception of the denoised recordings.

The paper is organized as follows. In Section II, the definition and characteristics/properties of the $\alpha-$stable distribution are described, while in Section III the $\alpha-$stable model and the inference of $\alpha-$stable model parameters is formulated. Signal modelling is presented in Section IV. In Section V, the experimental results are discussed. Conclusions are drawn in Section VI.

II. $\alpha-$Stable Distribution

A random variable (RV) X is drawn from a stable law distribution $f_{\gamma,\delta}(\alpha,\beta)$ iff its characteristic function is given by [6]:

$$\phi(\omega) = \exp\left(\gamma\, \psi_{\alpha,\beta}(\omega) + j\delta\omega\right) \tag{1}$$

where

$$\psi_{\alpha,\beta}(\omega) = \begin{cases} -|\omega|^{\alpha}[1 - j\,\mathrm{sign}\,(\omega)\beta \tan\frac{\pi\alpha}{2}], & \alpha \neq 1 \\ -|\omega|^{\alpha}[1 + j\,\mathrm{sign}\,(\omega)\beta \log|\omega|], & \alpha = 1 \end{cases} \tag{2}$$

and $-\infty < \delta < \infty$, $\gamma > 0$, $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$. Thus, a stable distribution is completely determined by four parameters: 1) the characteristic exponent α, 2) the index of skewness β, 3) the *scale* parameter γ, also called *dispersion*, and 4) the *location* parameter δ. A stable distribution with a characteristic exponent α is called $\alpha-$*stable*. The characteristic exponent α is a shape parameter, which measures the "thickness" of the tails of the density function and admits any value in the interval $0 < \alpha \leq 2$. If a stable RV is observed, the larger the value of α, the less likely is its realizations to deviate far from its central location. A small value of α implies considerable probability mass in the tails of the distribution. The index of skewness

β, which admits values in the interval $[-1, 1]$, determines the degree and sign of asymmetry. When $\beta = 0$, the distribution is symmetric about the center δ. Symmetric stable distributions with characteristic exponent α are called *symmetric* $\alpha-$*stable* (*SaS*). If $\alpha \neq 1$, the cases $\beta > 0$ and $\beta < 0$ correspond to left-skewness and right-skewness, respectively. The direction of skewness is reversed, if $\alpha = 1$ [2].

The notations $S(\alpha, \beta, \gamma, \delta)$ or $f_{\alpha,\beta}(\gamma, \delta)$ are often used to denote a stable distribution with parameters α, β, γ and δ. The PDF of a stable RV exists and is continuous, but it is not known in closed-form except the following three cases: 1) the Gaussian distribution $S(2, 0, \gamma, \delta) = \mathcal{N}(\delta, 2\gamma^2)$, i.e. a normal PDF with mean δ and variance $2\gamma^2$. 2) the Cauchy distribution $S(1, 0, \gamma, \delta)$ and 3) the Lévy distribution $S(0.5, 1, \gamma, \delta)$. For all the other cases, several estimation procedures for the PDF exist that rely on moment estimates or other sample statistics [7], [17]. The SaS distribution is represented as a SMIN [18] by exploiting the following product property of the symmetric $\alpha-$stable distribution [6], [14]:

Let $\mathbf{X}$ *and* $\mathbf{Y} > 0$ *be independent RVs with* $\mathbf{X} \sim f_{\alpha_1,0}(\sigma, 0)$ *and* $\mathbf{Y} \sim f_{\alpha_2,1}\left((\cos\frac{\pi\alpha_2}{2})^{1/\alpha_2}, 0\right)$*, then* $\mathbf{X}\mathbf{Y}^{1/\alpha_1} \sim f_{\alpha_1 \cdot \alpha_2,0}(\sigma, 0)$.

That is, let e_i be independent identically distributed (i.i.d) RVs drawn from symmetric $\alpha-$stable distribution with scale parameter γ and location parameter δ: $e_i \sim f_{\alpha,0}(\gamma, \delta)$. According to the product property, an equivalent representation exists where e_i is Gaussian conditionally on the auxiliary positive stable random variable ρ_i [14]:

$$e_i \sim \mathcal{N}(\delta, \rho_i\gamma^2) \tag{3}$$

$$\rho_i \sim f_{\alpha/2,1}\left(2\left(\cos\frac{\pi\alpha}{4}\right)^{2/\alpha}, 0\right). \tag{4}$$

III. $\alpha-$STABLE MODEL PARAMETER ESTIMATION

Following the Bayesian paradigm, the unknown quantities $\boldsymbol{\theta} = \{\alpha, \gamma, \delta\}$ can be inferred from the known data e:

$$p(\boldsymbol{\theta}|e) = p(\alpha, \gamma, \delta|e) \propto p(e|\alpha, \gamma, \delta)\, p(\alpha, \gamma, \delta). \tag{5}$$

Allowing the prior distribution $p(\alpha, \gamma, \delta)$ to depend on hyperpriors with their corresponding hyperparameters $\boldsymbol{\theta'} = \{\alpha', a_0, b_0, m_\delta, \sigma_\delta\}$ we get:

$$p(\boldsymbol{\theta}, \boldsymbol{\theta'}, e) = p(e|\boldsymbol{\theta}, \boldsymbol{\theta'})p(\boldsymbol{\theta}|\boldsymbol{\theta'})p(\boldsymbol{\theta'}). \tag{6}$$

The SaS graphical model depicted in Fig. 1, shows the conditional dependence structure between the parameters of the hyperparameters. In order to estimate the unknown parameters

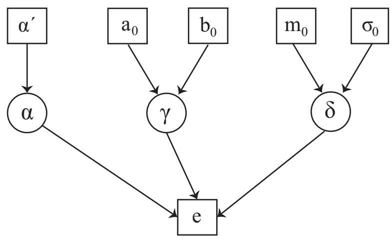


Fig. 1. Graphical model representation for the SaS model. Circles denote unknown variables, rectangles represent hyperparameters, while arrows denote the conditional dependence between variables.

of the SaS model, we sample from the posterior distribution of the parameters $\boldsymbol{\theta} = \{\alpha, \gamma, \delta\}$ using *MCMC* methods. To achieve this, conjugate priors for the unknown dispersion, location, and shape are considered in order to derive analytical expressions of the corresponding posterior distributions. The posterior distributions that are then sampled by means of Gibbs sampling in a straightforward way.

A. Prior distributions

Since the likelihood of the SaS model is Gaussian, conjugate priors are chosen for the dispersion and the location parameters in order to obtain a closed-form expression for the posterior [14]. The conjugate prior for the dispersion parameter γ is inverse Gamma with distribution:

$$p(\gamma^2|a_0, b_0) = \mathcal{IG}\left(a_0 + \frac{N}{2}, \frac{1}{2}\sum_{i=1}^{N}\frac{(y_i - \delta)^2}{\rho_i} + b_0\right). \tag{7}$$

The conjugate prior for the location parameter δ is Gaussian $\mathcal{N}(\delta|m_\delta, \sigma_\delta^{-1})$, i.e.

$$p(\delta|m_\delta, \sigma_\delta^{-1}) = \frac{1}{\sqrt{2\pi\sigma_\delta^{-1}}}\exp\left\{-\frac{(\delta - m_\delta)^2}{2\sigma_\delta^{-2}}\right\}. \tag{8}$$

For the parameter α, the uniform distribution is assumed in its support $\alpha \in (0, 2]$:

$$p(\alpha|\alpha') = \frac{1}{\alpha'} = \frac{1}{2},\ 0 < \alpha \leq 2. \tag{9}$$

B. MCMC Inference

Having chosen the prior distributions for the unknown parameters of the SaS model, the following MCMC scheme is used in order to sample from the corresponding posterior distributions:

1) *Updating the parameters* γ *and* δ *using Gibbs sampling.* The Gibbs sampler consists a standard MCMC technique that samples iteratively with replacement from the distribution of each parameter conditioned upon the others [19]. The samples obtained from the posterior distribution $p(\boldsymbol{\theta}|e)$ are then used to estimate the complete posterior density distribution.
   The conditional posterior distribution for the location parameter δ that has a Gaussian conjugate prior is given by [14]:

$$\mathcal{N}\left(\frac{\frac{1}{\gamma^2}\sum_{i=1}^{N}\frac{e_i}{\rho_i} + \sigma_\delta m_\delta}{\frac{1}{\gamma^2}\sum_{i=1}^{N}\frac{1}{\rho_i} + \sigma_\delta}, \frac{1}{\frac{1}{\gamma^2}\sum_{i=1}^{N}\frac{1}{\rho_i} + \sigma_\delta}\right). \tag{10}$$

   The full conditional for γ^2 that has an inverse Gamma conjugate prior is the following inverse Gamma distribution [14]:

$$\mathcal{IG}\left(a_0 + \frac{N}{2}, \frac{1}{2}\sum_{i=1}^{N}(e_i - \delta)^2 + b_0\right). \tag{11}$$

2) *Updating the parameter* α *using Metropolis sampling.* The Metropolis - Hastings algorithm (M-H) [20], [21] is used to estimate the parameter α, since the corresponding conditional distribution for α is unknown.

a) At each iteration t a candidate point α^{new} for α is generated from a proposal symmetric distribution $q(\cdot|\cdot)$. That is, $\alpha^{new} \sim q(\alpha^{new}|\alpha^{(t)})$.
b) $\mathcal{U}$ is generated from a uniform $(0, 1)$ distribution.
c) If $\mathcal{U} \leq A(\alpha^{new}|\alpha^{(t)})$ then α^{new} is accepted (i.e., $\alpha^{(t+1)} = \alpha^{new}$), otherwise α^{new} is rejected (i.e., $\alpha^{(t+1)} = \alpha^{(t)}$). That is, the candidate point α^{new} is accepted with probability $\min\{1, A\}$. Given that the proposal distribution $q(\cdot|\cdot)$ is symmetrical (i.e., $q(\alpha^{new}|\alpha^{(t)})$=$q(\alpha^{(t)}|\alpha^{new})$) and a uniform prior $p(\alpha)$ is considered (i.e., (9)), the acceptance/rejection ratio A is given by:

$$A = \min\left\{1, \frac{\prod_{i=1}^{N} p(e_i|\alpha^{new}, 0, \gamma, \delta)}{\prod_{i=1}^{N} p(e_i|\alpha^{(t)}, 0, \gamma, \delta)}\right\} \quad (12)$$

where $p(e_i|\alpha^{new}, 0, \gamma, \delta)$ and $p(e_i|\alpha^{(t)}, 0, \gamma, \delta)$ are calculated for the probability density function as in [6], [22][1].

3) *Estimating the auxiliary variable ρ_i using rejection sampling.*
Rejection sampling is used to sample from the posterior distribution $p(\rho)$

$$p(\rho_i|e_i, \gamma, \delta) \propto \mathcal{N}(e_i|\delta, \rho_i\gamma^2)\cdot \cdot f_{a/2,1}\left(\rho_i\middle|2\left(\cos\frac{\pi\alpha}{4}\right)^{2/\alpha}, 0\right). \quad (13)$$

It is readily seen that the likelihood forms a valid rejection function as it is bounded from above:

$$p(e_i|\delta, \rho_i\gamma^2) \leq \frac{1}{\sqrt{2\pi}|e_i - \delta|}\exp\left(-\frac{1}{2}\right). \quad (14)$$

Hence, the following rejection sampler can be used to draw samples from ρ_i [14]:

i. Samples are drawn from the positive stable distribution $\rho_i \sim f_{a/2,1}\left(2\left(\cos\frac{\pi\alpha}{4}\right)^{2/\alpha}, 0\right)$.
ii. Samples are drawn from the following uniform distribution $u_i \sim \mathcal{U}\left(0, \frac{1}{\sqrt{2\pi}|e_i-\delta|}\exp\left(-\frac{1}{2}\right)\right)$.
iii. If $u_i > p(e_i|\delta, \rho_i\gamma^2)$ goto step (i).

IV. Signal Model

The audio signal is modeled by three-layers associated to tones, transients, and noise [16]. Tones and transients are captured by decomposing the audio signal into two types of Modified Discrete Cosine Transform (MDCT) atoms [15], one having long frame length for the tones and one having short frame length for the transients, while noise is modeled as SaS noise. That is, the observed signal model is given by:

$$x = \sum_{k=1}^{N} \tilde{s}_{1,k}\Phi_{1,k} + \sum_{k=1}^{N} \tilde{s}_{2,k}\Phi_{2,k} + e \quad (15)$$

where N is the number of samples, $\Phi_{1,k}$ and $\Phi_{2,k}$ are the MDCT bases and e is the noise term having SaS distribution with scale γ and location δ modeled as described in (3) and (4). For simplicity reasons, we assume that $\delta = 0$. The clean audio signal contains a limited number of frequencies. Thus, the two vectors $\tilde{s}_1 = [\tilde{s}_{1,1}, \ldots, \tilde{s}_{1,N}]^T$ and $\tilde{s}_2 = [\tilde{s}_{2,1}, \ldots, \tilde{s}_{2,N}]^T$ are sparse. To model the sparsity in coefficients $\tilde{s}_{i,k}$, indicator binary random variables $g_{i,k} \in \{0, 1\}$ are introduced. When $g_{i,k} = 1$, the corresponding coefficient $\tilde{s}_{i,k}$ has a normal distribution. If $g_{i,k} = 0$, $\tilde{s}_{i,k}$ is set to zero enforcing sparsity to this coefficient [16]. The parameters of the signal model are estimated by means of MCMC methods. Therefore, appropriate conjugate priors are chosen for the model parameters in order to come up with analytical expressions for the corresponding posterior distributions.

A. Prior Distributions

1) *Coefficient priors*: The hierarchical prior for the coefficients is given by [16]:

$$p(\tilde{s}_{i,k}) = (1 - g_{i,k})\delta_0(\tilde{s}_{i,k}) + g_{i,k}\mathcal{N}(\tilde{s}_{i,k}|0, v_{i.k}) \quad (16)$$

where $\delta_0(\cdot)$ is the Dirac delta function, and $v_{i,k}$ has a conjugate inverse Gamma prior described by $p(v_{i,k}) = \mathcal{IG}(v_{i,k}|a_i, h_{i,k})$ with parameters a_i and $h_{i,k}$. $h_{i,k}$ is a parametric frequency profile expressed for each frequency index $j = 1, \ldots, l_{frame_i}$ by a Butterworth low-pass filter with filter order ν_i, cut-off frequency ω_i, and gain η_i:

$$h_{i,k} = \frac{\eta_i}{1 + \left(1 + \frac{j-1}{\omega_i}\right)^{\nu_i}}, k = (j, n), \quad (17)$$

and $n = 1, 2, \ldots, n_{frame_i}$ being a frame index, with $l_{frame_i} \times n_{frame_i} = N$, $i = 1, 2$.

2) *Indicator variable priors*: The indicator variables of the first basis corresponding to tonal parts are given a horizontal prior structure, while the indicator variables of the second basis corresponding to transient parts are given a vertical structure. In more detail, the sequence of indicator variables for the first basis are modeled by a two-state first-order Markov chain with transition probabilities $P_{1,00}$ and $P_{1,11}$ considered equal for all frequency indices [16]. The initial distribution $\pi_1 = P(g_{1,(j,1)} = 1))$ is given by its stationary distribution, $\pi_1 = \frac{1-P_{1,00}}{2-P_{1,11}-P_{1,00}}$ and $(1 - \pi_1) = \frac{1-P_{1,11}}{2-P_{1,11}-P_{1,00}}$. The transition probabilities $P_{1,00}$ and $P_{1,11}$ are given Beta priors, $\mathcal{B}(P_{1,00}|a_{P_{1,00}}, b_{P_{1,00}})$ and $\mathcal{B}(P_{1,11}|a_{P_{1,11}}, b_{P_{1,11}})$ Similarly, for the second basis the corresponding transition probabilities $P_{2,00}$ and $P_{2,11}$ are considered equal for all frames and are given Beta priors, $\mathcal{B}(P_{2,00}|a_{P_{2,00}}, b_{P_{2,00}})$ and $\mathcal{B}(P_{2,11}|a_{P_{2,11}}, b_{P_{2,11}})$. The initial distribution $\pi_2 = P(g_{2,(1,n)} = 1))$ is learned given Beta prior $\mathcal{B}(\pi_2|a_{\pi_2}, b_{\pi_2})$.

3) *Gain parameter prior*: The gain parameter η_i of the filter described in eq. (17) is given a Gamma conjugate prior, $p(\eta_i|a_{\eta_i}, b_{\eta_i}) = \mathcal{G}(\eta_i|a_{\eta_i}, b_{\eta_i})$ [16].

B. MCMC Inference

The following MCMC scheme is used to sample from the posterior distribution of the parameters $\boldsymbol{\theta} = \{\tilde{s}_i, v_i, \eta_i, P_{i,00}, P_{i,11}\}_{i=1,2} \cup \{\pi_2, \rho_i\gamma^2\}$ [16].

1) *Alternate sampling of $(g_1, \tilde{s}_1)$ and $(g_2, \tilde{s}_2)$.*

[1] http://www.mathworks.com/matlabcentral/fileexchange/37514-stbl-alpha-stable-distributions-for-matlab/content/STBL_CODE/stblpdf.m

The parameters $(g_1, \tilde{s}_1)$ and $(g_2, \tilde{s}_2)$ are alternatively sampled one after the other. The likelihood of the observation x is written as follows

$$p(x|\boldsymbol{\theta}) \sim \exp\left(-\frac{1}{2\gamma^2}\left\|\Sigma_\rho(x - \Phi_1\tilde{s}_1 - \Phi_2\tilde{s}_2)\right\|^2\right) \quad (18)$$

where Σ_ρ is a diagonal matrix with diagonal elements $[1/\sqrt{\rho_1}, \ldots, 1/\sqrt{\rho_N}]$ and $\|\cdot\|$ is the Frobenius norm.

2) *Updating of $(g_i, \tilde{s}_i)$ using Gibbs sampling.*
Let $\tilde{x}_{i|-i}$ be either $\tilde{x}_{1|2}$ or $\tilde{x}_{2|1}$. A Gibbs sampler is implemented that samples $(\tilde{s}_{i,k}, g_{i,k})$ jointly. Denoting by $g_{i,-k}$ the set $\{g_{i,1}, \ldots, g_{i,k-1}, g_{i,k+1}, \ldots, g_{i,N}\}$ and θ_{g_i} the set of Markov probabilities for g_i, (a) $g_{i,k}^{(l)}$ is sampled from $p(g_{i,k}^{(l)}|g_{i,-k}, \theta_{g_i}, v_i, \rho_i\gamma^2, \tilde{x}_{i|-i,k})$ and (b) $\tilde{s}_{i,k}^{(l)}$ is sampled from $p(\tilde{s}_{i,k}|g_{i,k}^{(l)}, v_i, \rho_i\gamma^2\tilde{x}_{i|-i,k})$. An hypothesis testing problem is set to estimate the first posterior probability for $g_{i,k}$ [23]:

$$H_0 \quad : \quad g_{i,k} = 1 \Longleftrightarrow \tilde{x}_{-i,k} = \tilde{s}_{i,k} + \tilde{e}_{i|-i,k} \quad (19)$$
$$H_1 \quad : \quad g_{i,k} = 0 \Longleftrightarrow \tilde{x}_{-i,k} = \tilde{e}_{i|-i,k}. \quad (20)$$

The following probabilities are thus used to draw values for $g_{i,k}$:

$$p(g_{i,k} = 0|g_{i,-k}, \theta_{g_i}, v_{i,k}, \rho_i\gamma^2, \tilde{x}_{i|-i,k}) = \frac{1}{1 + \tau_{i,k}} \quad (21)$$
$$p(g_{i,k} = 1|g_{i,-k}, \theta_{g_i}, v_{i,k}, \rho_i\gamma^2, \tilde{x}_{i|-i,k}) = \frac{\tau_{i,k}}{1 + \tau_{i,k}} \quad (22)$$

where

$$\begin{aligned}\tau_{i,k} &= \frac{p(g_{i,k} = 1|g_{i,-k}, \theta_{g_i}, v_{i,k}, \rho_i\gamma^2, \tilde{x}_{i|-i,k})}{p(g_{i,k} = 0|g_{i,-k}, \theta_{g_i}, v_{i,k}, \rho_i\gamma^2, \tilde{x}_{i|-i,k})} \\ &= \sqrt{\frac{\rho_i\gamma^2}{\rho_i\gamma^2 + v_{i,k}}}\exp\left(\frac{\tilde{x}_{i|-i,k}^2 v_{i,k}}{2\rho_i\gamma^2(\rho_i\gamma^2 + v_{i,k})}\right) \\ &\quad \times \frac{p(g_{i,k} = 1|g_{i,-k}, \theta_{g_i})}{p(g_{i,k} = 1|g_{i,-k}, \theta_{g_i})}. \end{aligned} \quad (23)$$

The posterior distribution for $\tilde{s}_{i,k}$ is given by

$$\begin{aligned}p(\tilde{s}_{i,k}|g_{i,k}, v_i, \rho_i\gamma^2, \tilde{x}_{i|-i,k}) &= (1 - g_{i,k})\delta_0(\tilde{s}_{i,k}) \\ &+ g_{i,k}\mathcal{N}(\tilde{s}_{i,k}|\mu_{\tilde{s}_{i,k}, \sigma^2_{\tilde{s}_{i,k}}}) \end{aligned} \quad (24)$$

where $\sigma^2_{\tilde{s}_{i,k}} = (1/\rho_i\gamma^2 + 1/v_{i,k})^{-1}$ and $\mu_{\tilde{s}_{i,k}} = (\sigma^2_{\tilde{s}_{i,k}}/\rho_i\gamma^2)\tilde{x}_{i|-i,k}$.

3) *Updating of v_i using Gibbs sampling.*
The conditional posterior distribution of $v_{i,k}$ is given by [16]

$$\begin{aligned}p(v_{i,k}|g_{i,k}, \tilde{s}_{i,k}, h_{i,k}) &= (1 - g_{i,k})\,\mathcal{IG}(v_{i,k}|a_i, h_{i,k}) \\ &+ g_{i,k}\,\mathcal{IG}\left(v_{i,k}\middle|\frac{1}{2} + a_i, \frac{\tilde{s}_{i,k}^2}{2} + h_{i,k}\right). \end{aligned} \quad (25)$$

4) *Updating of $\rho_i\gamma^2$ using Gibbs sampling.*

$$\begin{aligned}p(\rho_i\gamma^2|\tilde{s}_1, \tilde{s}_2, x) = \mathcal{IG}\Bigg(\rho_i\gamma^2\Bigg| & a_{\rho_i\gamma^2} + \frac{N}{2}, \\ & b_{\rho_i\gamma^2} + \frac{\|\Sigma_\rho(x - \Phi_1\tilde{s}_1 - \Phi_2\tilde{s}_2)\|^2}{2}\Bigg). \end{aligned} \quad (26)$$

5) *Updating of η_i using Gibbs sampling.*
The full posterior distribution of the gain parameter is given by

$$p(\eta_i|v_i) = \mathcal{G}\left(\eta_i\middle|Na_i + a_{\eta_i}, \sum_k \frac{1}{1 + \left(\frac{j-1}{\omega_i}\right)^{\nu_i} v_{i,k}} + b_{\eta_i}\right). \quad (27)$$

6) *Updating of $P_{i,00}$, $P_{i,11}$ and π_2.*

The posterior distributions of $P_{i,00}$, $P_{i,11}$ and π_2 are estimated by means of Metropolis-Hastings algorithm as described in [16] with corresponding proposed Beta distributions.

V. Experimental Results

A. Datasets and Parameters

8 musical excerpts ($\simeq 24s$ long each) from Greek folk songs recorded in outdoor festivities were used. Accordingly, the recordings are noisy. In all excerpts, a clarinet and a drum are playing. The songs were sampled at $44.1kHz$ resulting in $T = 2^{20} = 1048576$ samples for each song.

The algorithm described in Section IV was tested with the following parameter values: (a) The frame length for the tonals and the transients was respectively set to $l_{frame1} = 1024$ and $l_{frame_2} = 128$, respectively. The corresponding numbers of frames are thus $n_{frame_1} = 1024$ and $n_{frame_2} = 8192$ frames. (b) The filter-order ν_i in (17) was set to $\nu_1 = 6$ for the tonals and $\nu_2 = 4$ for the transients, respectively. (c) The cut-off frequency ω_i in (17) was set to $\omega_i = l_{frame_i}/3$. (d) The hyperparameters of the priors for η_i and $\rho_i\gamma^2$ were chosen to yield Jeffreys non-informative distributions. (e) The hyperparameters $a_{P_{i,00}}$, $a_{P_{i,11}}$ were set to values 50 and 1, respectively, weighing more heavily the values between 0.8 and 1. (f) The hyperparameters a_{π_2}, b_{π_2} for capturing the transients were set to 1 and 5000, respectively.

The Gibbs samplers described in Sections III and IV were run for 500 iterations with a burn-in period of 400 iterations. The estimate of the clean signal was constructed by $s^{(MMSE)} = \Phi_1\tilde{s}_1^{(MMSE)} + \Phi_2\tilde{s}_2^{(MMSE)}$, where $MMSE$ stands for the Minimum Mean Square Error estimates of the parameters, which were computed by averaging the values of $\tilde{s}_1$ and $\tilde{s}_2$ in the last 100 iterations of the sampler.

B. Performance

In order to measure the performance of a denoising algorithm, the overall output Noise Index (NI) defined as

$$NI_{db} = 20\log_{10}\frac{\|x\|}{\|x - s^{(MMSE)}\|} \quad (28)$$

is measured. Since NI expresses the ratio of the original noisy signal to the estimated noise, the smaller NI values imply a higher noise power removal and thus better denoising performance. The output NI values measured for the algorithm developed in Section IV, when α-stable noise residual is assumed in (15), are listed in Table I for several musical excerpts. In the same table, the output NI values measured for the original algorithm proposed in [16], that resorts to Gaussian noise residuals, are included.

TABLE I. OUTPUT NI VALUES OBTAINED BY THE PROPOSED ALGORITHM FOR α-STABLE NOISE NOISE RESIDUAL AND THE ALGORITHM IN [16] FOR GAUSSIAN WHITE NOISE RESIDUAL.

Index	*Song*	*NI*	
		α-stable noise	*Gaussian white noise*
1	Kalonixtia	15.2	49.2
2	Kastoriano syrto	15.3	45.1
3	To endika skropio	16.2	47.7
4	Paulos Milas	16.0	49.5
5	Sirto tou Panagioth	16.5	47.8
6	Kobo mia glara	15.3	47.1
7	Loukas	16.7	53.2
8	Mana mou ta louloudia sou	15.5	47.7

As can be seen in Table I, the assumption for an α-stable noise residual in (15) and the modifications due to this assumption in the framework proposed in [16] yields better denoising than for the assumption of a Gaussian white-noise residual. Furthermore, this fact can be verified by listening to the denoised musical excerpts[2]. When a Gaussian white noise residual is assumed, the processed audio files still contain a considerable amount of the background environmental noise. When an α-stable noise residual is assumed, the recordings are free from the background environmental noise, but they contain some new artifacts. In Fig. 2, the significance maps are depicted, when Greek folk song *2 (Kastoriano syrto)* is processed by the proposed algorithm that resorts to α-stable noise residual (a1-a2) and the algorithm in [16] that resorts to a Gaussian noise residual (b1-b2). The significance maps show the MMSE values of the indicator variables g_1 and g_2 of noise corruption for the tonals and the transients, respectively. The values range from 0 (white) to 1 (black). By comparing Fig. 2(a1) and Fig. 2(b1), it is seen that the proposed variant for the tonal layer yields similar results with the original algorithm in [16]. However, the performance of the two algorithms significantly differs for the transient layer. Indeed, the transient layer contains more artifacts, when a Gaussian noise residual is assumed as can be seen in Fig. 2(b2) than when an α-stable noise residual is assumed in the proposed variant of the algorithm in [16] (Fig. 2(a2)). The waveforms of the noisy and the filtered audio signal are plotted in Fig. 3.

The MCMC inference for the $\alpha-$stable parameters is shown in Fig. 4, where the values of the characteristic exponent α, the squared scale parameter γ^2, and the estimated standard deviation $\sqrt{\rho_i}\gamma$ of the $\alpha-$stable noise residual are depicted for each iteration of the Gibbs sampler. The MMSE estimates obtained by averaging the parameters in the last 100 iterations are: $\alpha \simeq 0.78$, $\gamma^2 \simeq 0.01$, and $\sqrt{\rho_i}\gamma \simeq 2$. Similarly, the MCMC inference for the remaining parameters of the signal model within the $\alpha-$stable noise is depicted in Fig. 5, where the Markov transition probabilities for the tonals and the transients $(P_{1,00}, P_{1,11})$, the gain parameter for the tonals and the transients (η_1, η_2), and the Markov initial probabilities for the transients (π_2) are plotted versus sampler iterations. The plots for the parameter values of the signal model in the original algorithm in [16] do not differ significantly from the plots shown in Fig. 5 with the exception of π_2 that converges slower when a Gaussian noise residual is assumed (Fig. 6) than when an $\alpha-$stable noise residual is considered (Fig. 5c).

All the aforementioned experiments were run on a Mac with a Core 2 Duo processor running at 2.4 GHz having 4 Gb RAM. On average, it took 314 min for the signal model with α-stable noise residual and 75 min for the signal model with Gaussian white noise residual to process a 24s long excerpt.

[2]Sample sound files are available at https://www.dropbox.com/sh/23iy9v52lbdl072/o32Znsm9oL

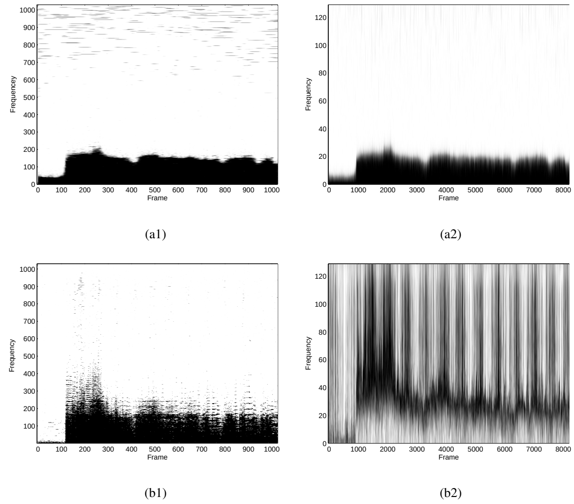


Fig. 2. Significance maps of the selected coefficients in Φ_1 and Φ_2 bases for the musical excerpt from Greek song *2 (Kastoriano syrto)*. The maps show the MMSE estimates of the noise indicator variables g_1 and g_2 when: (a1)-(a2) $\alpha-$stable noise residual and (b1)-(b2) Gaussian white noise residual is assumed in (15). The values range from 0 (white) to 1 (black).

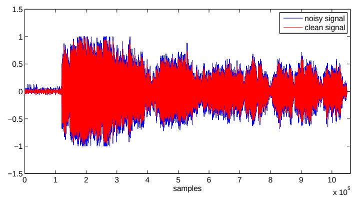


Fig. 3. Waveforms of the noisy and the estimated clean signal for the audio file *2 (Kastoriano syrto)*.

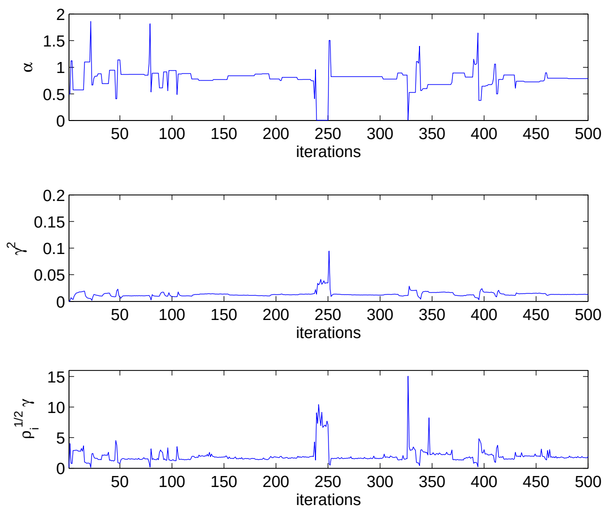
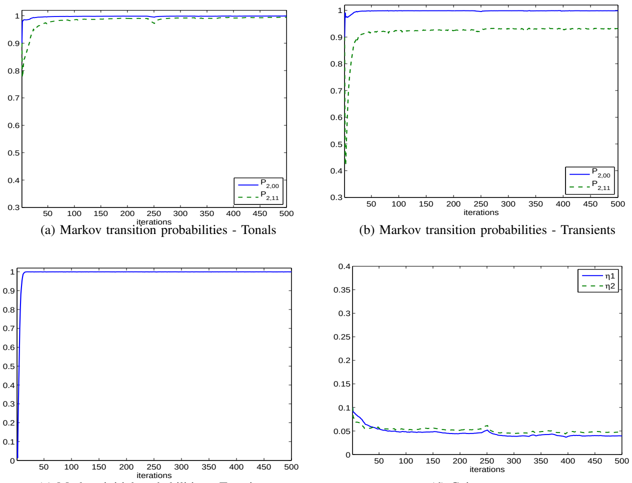


Fig. 4. Sampled values of the α-stable parameters: characteristic exponent α, square of scale parameter γ^2, and the standard deviation of the α-stable noise $\sqrt{\rho_i}\gamma$.



(a) Markov transition probabilities - Tonals
(b) Markov transition probabilities - Transients
(c) Markov initial probabilities - Transients
(d) Gain parameters

Fig. 5. Sampled values of (a) $P_{1,00}, P_{1,11}$, (b) $P_{2,00}, P_{2,11}$, (c) π_2, and (d) η_1, η_2 when an $\alpha-$stable noise is assumed.

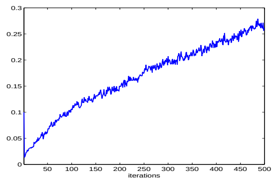


Fig. 6. Sampled values of π_2 when a Gaussian white noise is assumed.

The additional effort is due to the estimation of the α-stable parameters, with the estimation of ρ_i being the most time consuming.

VI. Conclusion

A musical audio denoising technique has been studied where the music signal is modeled in the frequency domain by two MDCT bases having indicator variables with structured priors and the residual noise is modeled by means of an α-stable distribution. The technique is formulated in a Bayesian setting and MCMC inference is used to estimate all the necessary model parameters. Preliminary results on musical excerpts from raw Greek folk songs recorded in outdoor festivities demonstrate that the α-stable noise assumption within the framework proposed in [16] is more suitable than the Gaussian white noise one. Testing the proposed technique in longer musical excerpts and examining the effect of model parameter initialization in more detail could be topics of future research.

Acknowledgment

This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operation Program "Education and Life-long Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALIS-UOA-ERASITECHNIS MIS 375435.

References

[1] I. Pitas and A. N. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications*. Dordrecht, The Netherlands: Kluwer Publishers, 1989.

[2] M. Shao and C. Nikias, "Signal processing with Fractional Lower Order Moments: Stable processes and their applications," *Proceeding of the IEEE*, vol. 81, no. 7, pp. 986–1010, July 1993.

[3] G. R. Arce, *Nonlinear Signal Processing*. Hoboken, NJ, USA: J. Wiley & Sons, 2005.

[4] T. Lemke and S. J. Godsill, "Linear Gaussian computations for near-exact Bayesian Monte Carlo inference in skewed $\alpha-$stable time series models," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 2012, pp. 3737–3740.

[5] V. M. Zolotarev, "One-dimensional stable distributions," *Translations of Mathematical Monographs, American Mathematical Society*, 1986.

[6] G. Samorodnitsky and M. Taqqu, *Stable non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. New York:Chapman and Hall, 1994.

[7] J. P. Nolan, *Stable Distributions: Models for Heavy-Tailed Data*. Birkhäuser, 2007.

[8] D. J. Buckle, "Bayesian inference for stable distributions," in *Journal of the American Statistical Association*, 1995, pp. 605–613.

[9] S. J. Godsill, "MCMC and EM-based methods for inference in heavy-tailed processes with alpha-stable innovations," in *Proc. IEEE Signal Processing Workshop on Higher-Order Statistics*, June 1999, pp. 228 – 232.

[10] S. J. Godsill and E. Kuruoğlu, "Bayesian inference for time series with heavy-tailed symmetric alpha-stable noise processes," in *Proc. Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, June 1999.

[11] E. Tsionas, "Monte Carlo inference in econometric models with symmetric stable disturbances," *Journal of Econometrics*, vol. 88, pp. 365–401, 1999.

[12] M. J. Lombardi, "Bayesian inference for $\alpha-$stable distributions: A random walk MCMC approach," *Computational Statistics and Data Analysis*, vol. 51, 2007.

[13] E. Kuruoğlu, "Analytical representation for positive $\alpha-$stable densities," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Hong Kong, 2003.

[14] D. Salas-Gonzalez, E. E. Kuruoglou, and D. P. Ruiz, "Modelling with mixture of symmetric stable distributions using Gibbs sampling," *Signal Processing*, vol. 90, no. 3, pp. 774–783, March 2010.

[15] H. S. Malvar, "Lapped transforms for efficient transform/subband coding," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 38, no. 6, pp. 969–978, June 1990.

[16] C. Fevotte, B. Torresani, l. Daudet, and S. Godsill, "Sparse linear regression with structured priors and application to denoising of musical audio," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 16, no. 1, pp. 174–185, January 2008.

[17] C. L. Nikias and M. Shao, *Signal Processing with α-Stable Distributions*. John Wiley and Sons, New York, 1995.

[18] M. J. Lombardi and S. J. Godsill, "On-line Baeysian estimation of signals in symmetric α-stable noise," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 775–779, 2006.

[19] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[20] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[21] W. Hastings, "Monte Carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[22] J. P. Nolan, "Numerical calculation of stable densities and distribution functions," *Commun. Statist. - Stochastic Models*, vol. 13, 1997.

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