Musical Audio Denoising Assuming Symmetric α-Stable Noise

Nikoletta Bassiou **Constantine Kotropoulos**
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki 54124, GREECE
nbassiou@aiia.csd.auth.gr costas@aiia.csd.auth.gr

ABSTRACT

The representation of α-stable distributions as scale mixture of normals is exploited to model the noise in musical audio recordings. Markov Chain Monte Carlo inference is used to estimate the clean signal model and the α-stable noise model parameters in a sparse linear regression framework with structured priors. The musical audio recordings were processed both as a whole and in segments by using a sine-bell window for analysis and overlap-and-add reconstruction. Experiments on noisy Greek folk music excerpts demonstrate better denoising under the α-stable noise assumption than the Gaussian white noise one, when processing is performed in segments rather than in full recordings.

1. INTRODUCTION

Signals contaminated by outliers (e.g., impulsive noise) or corrupted by noise generated by an asymmetric probability density function (PDF) cannot be accurately modeled by Gaussian statistics [1, 2]. The α-stable distributions are more suitable to model the aforementioned phenomena due to their properties, such as infinite variance, skewness, and heavy tails [3, 4]. Among the α-stable distributions, the symmetric ones have been extensively studied within a Bayesian framework, since the PDF of α-stable distributions cannot be analytically described in general. In [5], a particular mathematical representation was exploited to infer the α-stable parameters using the Gibbs sampler. Monte Carlo Expectation-Maximization and Markov Chain Monte Carlo (MCMC) methods were introduced in [6], which were based on the representation of α-stable distributions as Scale Mixture of Normals (SMiN). The SMiN property was also exploited to model symmetric α-stable (SaS) disturbances with a Gibbs Metropolis sampler [7]. Recently, a random walk MCMC approach for Bayesian inference in stable distributions was introduced using a numerical approximation of the likelihood function [8]. An analytical approximation of the positive α-stable distribution based on a product of a Pearson and another positive stable random variable was proposed in [9]. Finally, a Poisson sum series representation for the SaS distribution was used to express the noise process in a conditionally Gaussian framework [10].

A growing body of research aims at extending sparsity paradigms in order to better capture the structure of signals [11]. For audio signals, structure is a consequence of basic acoustic laws describing resonant systems and impact sounds, implying that large classes of audio components are either sparse in the frequency domain and persistent in time or sparse in time and persistent in frequency [12]. Here, the signal is modeled by two Modified Discrete Cosine Transform (MDCT) bases, one describing the tonal parts of the signal and one describing its transient parts [13]. Sparsity is enforced in the expansion coefficients of each MDCT base by means of binary indicator variables with structured priors as in [13]. Alternatively, one could employ Gabor frames and develop sparse expansions enforcing an ℓ_1 regularization to the expansion coefficients [14]. In this paper, a SaS distribution models the noise in recordings of Greek folk songs performed in outdoor festivities. Experimental evidence is disclosed that demonstrates the validity of this assumption. Indeed, both the probability-probability ($P-P$) plots and the sampled value of the characteristic exponent in the SaS distribution indicate that the noise statistics deviate from Gaussian ones. By modeling the noise by a SaS distribution, the framework in [13], where a Gaussian white noise was assumed only, is generalized. A standard MCMC technique is used to estimate the signal and the α-stable noise parameters following similar lines to [8, 15]. Extending the preliminary work [16], here the musical audio recordings are processed both as a whole and in segments by using a sine-bell window for analysis and overlap-and-add reconstruction. The experimental results demonstrate a superior performance for the SaS noise assumption in the overlap-and-add reconstruction with respect to the power of the noise remaining after denoising and the acoustic perception of the processed music recordings.

The paper is organized as follows. In Section 2 the definition and the properties of the α-stable distribution are reviewed. Section 3 is devoted to signal modeling, while the inference of $\alpha-$stable model parameters is studied in Section 4. Experimental results are demonstrated in Section 5 and conclusions are drawn in Section 6.

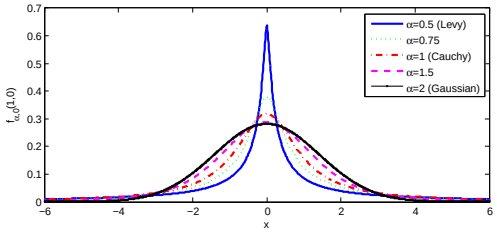


Figure 1. PDFs of normalized unitary dispersion SaS $f_{\alpha,0}(1,0)$ for various values of the tail constant α.

2. $\alpha-$STABLE DISTRIBUTION

A random variable (RV) X is drawn from a stable law distribution $f_{\alpha,\beta}(\gamma,\delta)$ iff its characteristic function is given by [3]:

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

where

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

with $-\infty < \delta < \infty$, $\gamma > 0$, $0 < \alpha \leq 2$, and $-1 \leq \beta \leq 1$. Accordingly, a stable distribution is completely determined by four parameters: 1) the characteristic exponent or tail constant 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 α-*stable*. The characteristic exponent α is a shape parameter, which measures the "thickness" of the tails of the density function. If a stable RV is observed, the larger the value of α, the less likely is to observe values of the RV, which are 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 determines the degree and sign of asymmetry. When $\beta = 0$, the distribution is symmetric about the center δ. SaS are symmetric stable distributions with characteristic exponent α. 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 $\alpha = 1$ [17].

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 stable random variables exist and are continuous, but they are not known in closed-form except the following three cases: 1) the Gaussian distribution $S(2,0,\gamma,\delta) = N(\delta, 2\gamma^2)$, 2) the Cauchy distribution $S(1,0,\gamma,\delta)$ and 3) the Lévy distribution $S(0.5,1,\gamma,\delta)$, which admit a closed-form PDF. For all the other cases, several estimation procedures for the PDF exist that rely on moment estimates or other sample statistics [4, 18]. Several SaS PDFs are plotted in Figure 1.

The symmetric α-stable distribution is represented as a scale of mixture of normals [19] by exploiting the following product property of the symmetric α-stable distribution [3, 15]: *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* $\mathbf{X}\mathbf{Y}^{1/\alpha_1} \sim f_{\alpha_1\cdot\alpha_2,0}(\sigma, 0)$.

3. SIGNAL MODEL

Let l_{frame} and n_{frame} denote the frame length and the number of frames. Their product equals the number of samples, N, in an audio recording. The observed audio signal is modeled by an underlying clean signal represented by two layers associated to tones or transients, and the corrupting noise [13]. Tones and transients are captured by decomposing the audio signal into two types of MDCT atoms [20], while noise is modeled as SaS noise. Let $\mathbf{\Phi}_1 = [\mathbf{\Phi}_{1,1}, \mathbf{\Phi}_{1,2}, \ldots, \mathbf{\Phi}_{1,N}] \in \mathbb{R}^{N\times N}$ be the MDCT base with long frame length l_{frame_1} representing the tonals and $\mathbf{\Phi}_2 = [\mathbf{\Phi}_{2,1}, \mathbf{\Phi}_{2,2}, \ldots, \mathbf{\Phi}_{2,N}] \in \mathbb{R}^{N\times N}$ be the MDCT base with short frame length l_{frame_2} representing the transients. For $i = 1, 2$, $N = l_{frame_i} \times n_{frame_i}$. The atoms of either basis $\mathbf{\Phi}_{i,k}$ are indexed by $k = 1, 2, \ldots, N$, such that $k = (n-1)\, l_{frame_i} + j$ where $j = 1, 2, \ldots, l_{frame_i}$ is a frequency index and $n = 1, 2, \ldots, n_{frame_i}$ is a frame index. Let also $\tilde{\mathbf{s}}_1$, $\tilde{\mathbf{s}}_2 \in \mathbb{R}^{N\times 1}$ be two coefficient vectors and $\mathbf{e} \in \mathbb{R}^{N\times 1}$ be the noise vector comprising independent identically distributed (i.i.d.) RVs drawn from a SaS distribution with characteristic exponent α, scale γ, and location parameter δ (i.e., $\mathbf{e} \sim f_{\alpha,0}(\gamma,\delta)$). Then, the observed signal model $\mathbf{x} \in \mathbb{R}^{N\times 1}$ is given by:

$$\mathbf{x} = \mathbf{\Phi}_1\, \tilde{\mathbf{s}}_1 + \mathbf{\Phi}_2\, \tilde{\mathbf{s}}_2 + \mathbf{e}. \tag{3}$$

That is, for $l = 1, 2, \ldots, N$, the lth element of the observed signal in the time domain is expressed as

$$x_l = \sum_{k=1}^{N} \Phi_{1,l,k}\, \tilde{s}_{1,k} + \sum_{k=1}^{N} \Phi_{2,l,k}\, \tilde{s}_{2,k} + e_l \tag{4}$$

where $\Phi_{i,l,k}$ is the lth element of $\mathbf{\Phi}_{i,k} \in \mathbb{R}^{N\times 1}$, $i = 1, 2$ and $k = 1, 2, \ldots, N$. The product property of the SaS distribution [3] suggests that the e_l are equivalently represented by a Gaussian RV conditionally independent on the auxiliary positive stable RV ρ_k [15]:

$$e_l \sim \mathcal{N}(\delta, \rho_l\gamma^2), \rho_l \sim f_{\alpha/2,1}\left(2\left(\cos\frac{\pi\alpha}{4}\right)^{2/\alpha}, 0\right). \tag{5}$$

The two coefficient vectors $\tilde{\mathbf{s}}_1$ and $\tilde{\mathbf{s}}_2$ are sparse, since the clean audio signal contains a limited number of frequencies. The sparsity in coefficients $\tilde{s}_{i,k}$, is modeled by means of indicator binary random variables $g_{i,k} \in \{0, 1\}$. When $g_{i,k} = 1$, the corresponding $\tilde{s}_{i,k}$ has a normal distribution. Otherwise, $\tilde{s}_{i,k}$ is set to zero enforcing sparsity to this coefficient [13]. The parameters of the underlying clean signal model are estimated by means of MCMC methods.

3.1 MCMC Inference

Let $\boldsymbol{\theta}$ collectively refer to the set of parameters to be sampled from their posterior distribution using the following MCMC scheme [13].

1. *Alternate sampling of* $(\mathbf{g}_1, \tilde{\mathbf{s}}_1)$ *and* $(\mathbf{g}_2, \tilde{\mathbf{s}}_2)$.
   The parameters $(\mathbf{g}_1, \tilde{\mathbf{s}}_1)$ and $(\mathbf{g}_2, \tilde{\mathbf{s}}_2)$ are sampled one

after the other in an alternating fashion. The likelihood of the observed audio signal $\mathbf{x}$ is written as follows

$$p(\mathbf{x}|\boldsymbol{\theta}) \sim \exp\left(-\frac{1}{2\gamma^2}\left\|\boldsymbol{\Sigma}_\rho(\mathbf{x}-\boldsymbol{\Phi}_1\tilde{\mathbf{s}}_1-\boldsymbol{\Phi_2}\tilde{\mathbf{s}}_2)\right\|^2\right) \tag{6}$$

where $\boldsymbol{\Sigma}_\rho$ is a diagonal matrix with diagonal elements $[1/\sqrt{\rho_1},\ldots,1/\sqrt{\rho_N}]$ and $\|\cdot\|$ is the ℓ_2 norm.

2. *Updating of* $(\mathbf{g}_i, \tilde{\mathbf{s}}_i)$ *using Gibbs sampling.*
Let $\tilde{\mathbf{x}}_{i|-i}$ be either $\tilde{\mathbf{x}}_{1|2} = \boldsymbol{\Phi}_1^\top(\mathbf{x}-\boldsymbol{\Phi}_2\tilde{\mathbf{s}}_2)$ or $\tilde{\mathbf{x}}_{2|1} = \boldsymbol{\Phi}_2^\top(\mathbf{x}-\boldsymbol{\Phi}_1\tilde{\mathbf{s}}_1)$, and $\tilde{\mathbf{e}}_i = \boldsymbol{\Phi}_i^\top\mathbf{e}$. $\tilde{s}_{i,k}$ is given a hierarchical prior described by $p(\tilde{s}_{i,k}) = (1-g_{i,k})\delta_0(\tilde{s}_{i,k}) + g_{i,k}\mathcal{N}(\tilde{s}_{i,k}|0,v_{i.k})$ with $\delta_0(\cdot)$ being the Dirac delta function, and $v_{i,k}$ having a conjugate inverse Gamma prior, $p(v_{i,k}) = \mathcal{IG}(v_{i,k}|a_i,h_{i,k})$. $h_{i,k}$ 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 [13]. Then, 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, $g_{i,k}$ at the tth iteration, $g_{i,k}^{(t)}$, is sampled from $p(g_{i,k}|g_{i,-k},\theta_{g_i}, v_i,\rho_l\gamma^2,\tilde{x}_{i|-i,k})$ and $\tilde{s}_{i,k}^{(t)}$ is sampled from $p(\tilde{s}_{i,k}|g_{i,k}^{(t)}, v_i,\rho_l\gamma^2,\tilde{x}_{i|-i,k})$. A hypothesis testing problem is set to estimate the first posterior probability for $g_{i,k}$ [21]:

$$H_1 \;:\; g_{i,k}=1 \Longleftrightarrow \tilde{x}_{i|-i,k} = \tilde{s}_{i,k}+\tilde{e}_{i,k} \tag{7}$$
$$H_0 \;:\; g_{i,k}=0 \Longleftrightarrow \tilde{x}_{i|-i,k} = \tilde{e}_{i,k}. \tag{8}$$

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

$$\begin{aligned} p(g_{i,k}=0|g_{i,-k},\theta_{g_i},v_{i,k},\rho_l\gamma^2,\tilde{x}_{i|-i,k}) &= \frac{1}{1+\tau_{i,k}} \\ p(g_{i,k}=1|g_{i,-k},\theta_{g_i},v_{i,k},\rho_l\gamma^2,\tilde{x}_{i|-i,k}) &= \frac{\tau_{i,k}}{1+\tau_{i,k}} \end{aligned} \tag{9}$$

where

$$\begin{aligned} \tau_{i,k} &= \sqrt{\frac{\rho_l\gamma^2}{\rho_l\gamma^2+v_{i,k}}}\exp\left(\frac{\tilde{x}_{i|-i,k}^2 v_{i,k}}{2\rho_l\gamma^2(\rho_l\gamma^2+v_{i,k})}\right) \\ &\quad\times\frac{p(g_{i,k}=1|g_{i,-k},\theta_{g_i})}{p(g_{i,k}=0|g_{i,-k},\theta_{g_i})}. \end{aligned} \tag{10}$$

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

$$\begin{aligned} p(\tilde{s}_{i,k}|g_{i,k},v_i,\rho_l\gamma^2,\tilde{x}_{i|-i,k}) &= (1-g_{i,k})\delta_0(\tilde{s}_{i,k}) \\ &\quad + g_{i,k}\mathcal{N}\left(\tilde{s}_{i,k}|\mu_{\tilde{s}_{i,k}}\sigma^2_{\tilde{s}_{i,k}}\right) \end{aligned} \tag{11}$$

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

3. *Updating of* v_i *using Gibbs sampling.*
The conditional posterior distribution of $v_{i,k}$ is given by $p(v_{i,k}|g_{i,k},\tilde{s}_{i,k},h_{i,k}) = (1-g_{i,k})\,\mathcal{IG}(v_{i,k}|a_i,h_{i,k}) + g_{i,k}\,\mathcal{IG}\left(v_{i,k}\left|\frac{1}{2}+a_i,\frac{\tilde{s}_{i,k}^2}{2}+h_{i,k}\right.\right)$ [13].

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

$$\begin{aligned} p(\rho_l\gamma^2|\tilde{\mathbf{s}}_1,\tilde{\mathbf{s}}_2,\mathbf{x}) &= \mathcal{IG}\left(\rho_l\gamma^2|a_{\rho_l\gamma^2}+N/2,\right. \\ &\quad \left. b_{\rho_l\gamma^2}+(\|\boldsymbol{\Sigma}_\rho(\mathbf{x}-\boldsymbol{\Phi}_1\tilde{\mathbf{s}}_1-\boldsymbol{\Phi}_2\tilde{\mathbf{s}}_2)\|^2)/2\right) \end{aligned} \tag{12}$$

5. *Updating of* η_i *using Gibbs sampling.*
The gain parameter η_i of the Butterworth filter is given a Gamma conjugate prior, $p(\eta_i|a_{\eta_i},b_{\eta_i}) = \mathcal{G}(\eta_i|a_{\eta_i},b_{\eta_i})$ [13].
The full posterior distribution of the gain parameter η_i is $p(\eta_i|v_i) = \mathcal{G}\left(\eta_i\left|Na_i+a_{\eta_i},\sum_k\frac{1}{1+\left(\frac{j-1}{\omega_i}\right)^{\nu_i}v_{i,k}}+b_{\eta_i}\right.\right)$ [13].

6. *Updating of* $P_{i,00}$, $P_{i,11}$, *and* π_2.
The indicator variables of the first basis corresponding to tonal parts are given a horizontal prior structure and 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 [13]. The initial distribution $\pi_1 = P(g_{1,(j,1)}=1))$ is its stationary distribution, $\pi_1 = \frac{1-P_{1,00}}{2-P_{1,11}-P_{1,00}}$. 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 $\mathcal{B}(P_{1,11}|a_{P_{1,11}},b_{P_{1,11}})$, respectively. The indicator variables of the second basis corresponding to transient parts are given a vertical structure. 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 $\mathcal{B}(P_{2,11}|a_{P_{2,11}},b_{P_{2,11}})$ as well. The initial distribution $\pi_2 = P(g_{2,(1,n)}=1)$ is learned given a Beta prior $\mathcal{B}(\pi_2|a_{\pi_2},b_{\pi_2})$.
The posterior distributions of $P_{i,00}$, $P_{i,11}$ and π_2 are estimated by means of the Metropolis-Hastings (M-H) algorithm as described in [13] with corresponding proposed Beta distributions.

4. SAS MODEL PARAMETER ESTIMATION

Similarly to the signal model, in order to estimate the unknown SaS parameters of the noise model (5), we sample from the posterior distribution of the parameters $\boldsymbol{\theta} = \{\alpha,\gamma,\delta\}$ using MCMC methods with appropriate conjugate priors chosen for the model parameters.

4.1 MCMC Inference

1. *Updating parameters* γ *and* δ *using Gibbs sampling.*
The conditional posterior distribution for the location parameter δ with a Gaussian conjugate prior [16] is: $\mathcal{N}\left(\frac{\frac{1}{\gamma^2}\sum_{l=1}^N\frac{e_l}{\rho_l}+\sigma_\delta m_\delta}{\frac{1}{\gamma^2}\sum_{l=1}^N\frac{1}{\rho_l}+\sigma_\delta},\frac{1}{\frac{1}{\gamma^2}\sum_{l=1}^N\frac{1}{\rho_l}+\sigma_\delta}\right)$ [15].
The full conditional for γ^2, that has an inverse Gamma conjugate prior [16], is the inverse Gamma distribution $\mathcal{IG}\left(a_0+\frac{N}{2},\frac{1}{2}\sum_{l=1}^N(e_l-\delta)^2+b_0\right)$ [15].

2. *Updating the parameter* α *using Metropolis sampling.*
The M-H algorithm [22, 23] 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, otherwise α^{new} is rejected. That is, the candidate point α^{new} is accepted with probability $\min\{1, A\}$. Given that the proposal distribution $q(\cdot|\cdot)$ is symmetrical and considering a uniform prior, $p(\alpha|\alpha') = \frac{1}{\alpha'} = \frac{1}{2}$, $0 < \alpha \leq 2$, the acceptance/rejection ratio A is given by $A = \min\left\{1, \frac{\prod_{l=1}^{N} p(e_l|\alpha^{new},0,\gamma,\delta)}{\prod_{l=1}^{N} p(e_l|\alpha^{(t)},0,\gamma,\delta)}\right\}$ where $p(e_l|\alpha^{new}, 0, \gamma, \delta)$ and $p(e_l|\alpha^{(t)}, 0, \gamma, \delta)$ are calculated for the probability density function as in [3, 24] [1] .

3. *Estimating auxiliary variable ρ_l with rejection sampling.*

Rejection sampling is used to sample from the posterior distribution

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

The likelihood forms a valid rejection function as it is bounded from above $p(e_l|\delta, \rho_l\gamma^2) \leq \frac{\exp\left(-\frac{1}{2}\right)}{\sqrt{2\pi}|e_l-\delta|}$. Hence, the following rejection sampler can be used to draw samples from ρ_k [15]:

i. Samples are drawn from the positive stable distribution $\rho_l \sim f_{a/2,1}\left(2\left(\cos\frac{\pi\alpha}{4}\right)^{2/\alpha}, 0\right)$.

ii. Samples are drawn from the following uniform distribution $u_l \sim \mathcal{U}\left(0, \frac{1}{\sqrt{2\pi}|e_l-\delta|}\exp\left(-\frac{1}{2}\right)\right)$.

iii. If $u_l > p(e_l|\delta, \rho_l\gamma^2)$ go to step i.

5. EXPERIMENTAL RESULTS

4 noisy musical excerpts ($\simeq 48s$ long each) from Greek folk songs recorded in outdoor festivities were used. In all excerpts, a clarinet and a drum are playing. The songs were sampled at 44.1 kHz resulting in $T = 2^{21} = 2097152$ samples for each song. They were also segmented in 17 and 67 "superframes" with 131072 and 32768 samples each, respectively. In both cases, the superframes were overlapping by 1024 samples. A sine-bell window was used for analysis and overlap-and-add reconstruction of the full denoised signals. The denoising algorithm was tested for restoring the excerpts as a whole as well as restoring the superframes in every excerpt for the following parameter values: (a) $l_{frame1} = 1024$ and $l_{frame_2} = 128$, resulting in $n_{frame_1} = 2048$ and $n_{frame_2} = 16384$ frames, respectively. (b) The Butterworth filter parameters were respectively set to $\omega_i = l_{frame_i}/3$ and $\nu_1 = 6$ and $\nu_2 = 4$. (c) η_i and $\rho_l\gamma^2$ were chosen to yield Jeffreys non-informative distributions. (d) The hyperparameters for $P_{i,00}, P_{i,11}$ and π_2 were set to $a_{P_{i,00}} = 50$, $a_{P_{i,11}} = 1$, $a_{\pi_2} = 1$, and $b_{\pi_2} = 5000$. (e) The Gibbs samplers described in Sections 3 and 4 were run for 300 iterations with a burn-in period of 240 iterations. The clean signal was estimated by $\mathbf{s}^{(MMSE)} = \mathbf{\Phi}_1\ \tilde{\mathbf{s}}_1^{(MMSE)} + \mathbf{\Phi}_2\ \tilde{\mathbf{s}}_2^{(MMSE)}$, where $MMSE$ stands for the Minimum Mean Square Error estimates of $\tilde{\mathbf{s}}_1$ and $\tilde{\mathbf{s}}_2$, which were computed by averaging their values in the last 60 iterations of the sampler.

The performance of the denoising algorithm is measured by means of the overall output Noise Index (NI) [16], which expresses the ratio of the original noisy signal to the estimated noise, i.e.,

$$NI_{db} = 20\log_{10}\frac{\|\mathbf{x}\|^2}{\|\mathbf{x} - \mathbf{s}^{(MMSE)}\|^2} \quad (14)$$

The smaller NI value implies the higher noise power removal and thus a better denoising performance. The output NI values measured for the algorithm developed in Section 3, when α-stable noise residual is assumed in (3), are listed in Table 1 for the musical excerpts processed both as a whole as well as in segments using overlap-and-add reconstruction. In the same table, the output NI values measured for the original algorithm proposed in [13] that resorts to Gaussian noise residuals, are included. As can be seen in Table 1, the assumption for a SaS noise residual in (3) and the modifications made due to this assumption in the framework proposed in [13] yields better denoising than the assumption of a Gaussian white-noise residual. Especially, for the SaS noise residual assumption, the denoising performance is considerably improved when the musical excerpts are processed in segments and overlap-and-add reconstruction, while the denoising performance improvement when assuming a Gaussian white-noise residual is negligible.

The aforementioned conclusions are also 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 recording noise together with some new artifacts. When a SaS noise residual is assumed, the recordings are free from recording noise, but some cracks are inserted.

In Figure 2, the significance maps are depicted, when the fourth Greek folk song is processed by the proposed algorithm that resorts to SaS noise residual (a1-a2) and the algorithm in [13] that resorts to a Gaussian noise residual (b1-b2). By comparing Figure 2(a1) and Figure 2(b1), it is seen that the proposed variant for the tonal layer yields similar results with the original algorithm in [13]. However, the performance of the two algorithms significantly differs for the transient layer, since more artifacts are present, when a Gaussian noise residual is assumed (Figure 2 (b2)) than when a SaS stable noise residual is assumed in the proposed variant of the algorithm in [13] (Figure 2(a2)).

Spectrograms of a 6s long excerpt extracted from the 4th song are shown in Figure 3. The spectrogram of the raw recording is shown in Figure 3(a). The spectrograms of the

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

[2] https://www.dropbox.com/sh/jz65g0tgx5q05j5/e4vktfFvxl

Ind.	*Song*	SaS noise			Gaussian white noise		
		no oa	*oa_1*	*oa_2*	*no oa*	*oa_1*	*oa_2*
1	Kalonixtia (Good night)	35.0	26.6	29.8	48.5	48.4	48.2
2	Loukas (Luke)	39.0	26.5	29.7	51.7	50.8	50.7
3	To endika skorpio (Scatter at 11 o' clock)	31.7	27.2	31.5	49.2	48.9	49.1
4	Sirto Panagioti (Panagiotis' Syrtos)	38.8	26.4	29.1	47.1	47.3	47.4
5	Paulos Milas (Paulos Melas)	33.7	27.3	30.0	47.7	47.5	47.4

Table 1. Output NI values of the proposed algorithm for SaS noise residual and the algorithm in [13] for Gaussian white noise residual applied on the musical excerpts processed as a whole (no oa) and in superframes by means of overlap-and-add reconstruction (oa_1: 131072 samples long, and oa_2: 32768 samples long).

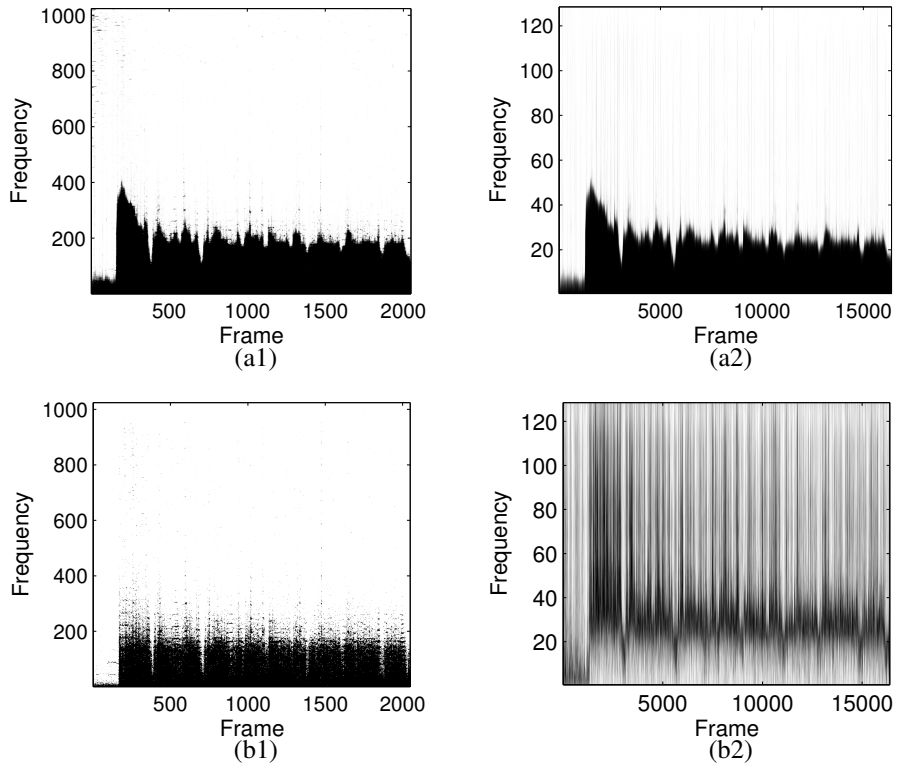
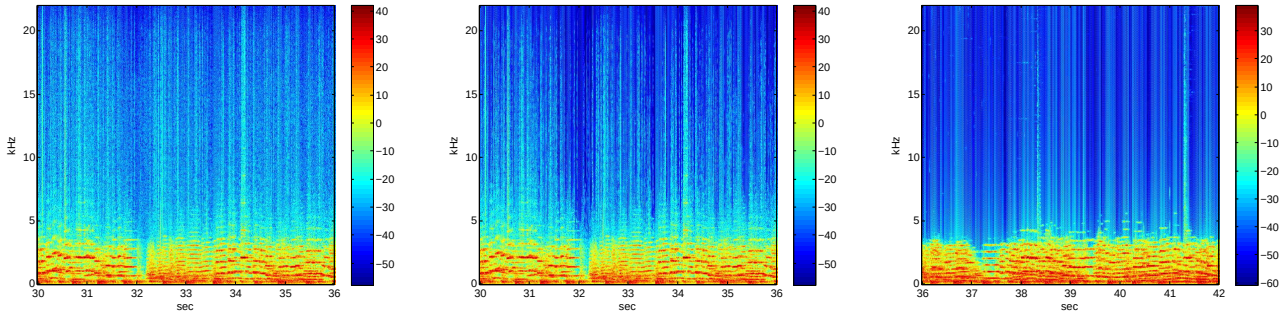


Figure 2. Significance maps of the selected coefficients in $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ bases for the musical excerpt 4. The maps show the MMSE estimates of the noise indicator variables g_1 and g_2 for: (a1)-(a2) SaS noise residual and (b1)-(b2) Gaussian white noise residual in (3). The values range from 0 (white) to 1 (black).

denoised recordings that were reconstructed by the overlap-and-add method, when either a Gaussian or a SaS noise residual is assumed are shown in Figures 3(b) and (c). In the reconstruction, 131072 samples long superframes were employed. The inspection of Figure 3(c) reveals the superior denoising performance when a SaS noise residual is assumed.

The MCMC inference for the SaS parameters is shown in Figure 4, where the values of the characteristic exponent α and the estimated standard deviation $\sqrt{\rho_l}\gamma$ of the SaS noise residual averaged across the last 60 iterations of the Gibbs sampler are depicted for each segment of the musical excerpt reconstructed by means of the overlap-and-add method. The corresponding mean values are: $\alpha \simeq 0.2$ and $\alpha \simeq 0.25$ for the overlap-and-add with 131072 and 32768 samples, respectively, and $\sqrt{\rho_l}\gamma \simeq 2.4$ and $\sqrt{\rho_l}\gamma \simeq 2.6$ for the overlap-and-add with 131072 and 32768 samples, respectively. The mean values for the stable parameter δ are of the order of 10^{-4} in all cases, as expected. The sampled values of the characteristic exponent indicate a strong deviation from the Gaussian statistics corresponding to $\alpha = 2$.

Furthermore, three PDFs were tested for modeling the noisy segments of recordings, namely the Gaussian, the Student-t, and the SaS. The $P-P$ plots were used for that purpose. Let $F()$ denote the cumulative density function associated to a model. For estimates of location and scale parameters, $\hat{\mu}$ and $\hat{\sigma}$, a $P-P$ plot is defined by the set of points $(\xi_l, F(\frac{x_{(l)}-\hat{\mu}}{\hat{\sigma}})$ with $l = 1, 2, \ldots, N$, where $\xi_l = \frac{l}{N+1}$ and $x_{(l)}$ are the observations arranged in ascending order of magnitude, i.e., $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(N)}$. A



(a) (b) (c)

Figure 3. Spectrograms of a 6s long excerpt from the 4th excerpt for the: (a) Original recording. (b) Denoised recording reconstructed by overlap-and-add, when a Gaussian noise residual is assumed and segments of 131072 samples were employed. (c) Denoised recording reconstructed by overlap-and-add, when a SaS noise residual is assumed and segments of 131072 samples were employed.

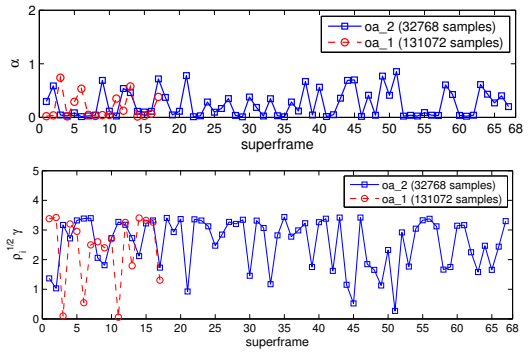


Figure 4. Sampled values of the characteristic exponent α and the standard deviation $\sqrt{\rho_l}\gamma$ of the SaS noise, averaged across the iterations of the Gibbs sampler for each superframe of the overlap-and-add method.

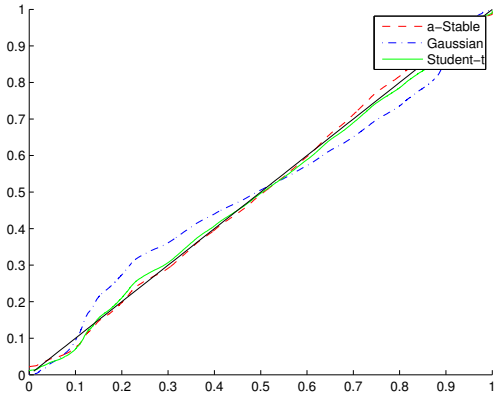


Figure 5. $P-P$ plot for noisy observations extracted from song 4 in Table 1 (Sirto Panagioti).

strong deviation of the $P-P$ plot from the main diagonal in the unit square indicates the the model assumed is incorrect (or the location and scale parameters are inaccurate). The $P-P$ plots from the aforementioned models are shown in Figure 5. It is seen that the $P-P$ plot for the SaS model lies much closer to the main diagonal than that of the Gaussian and the Student-t models.

All the experiments were run on a Mac Core 2 Duo running at 2.4 GHz with 8 GB RAM. On average, in the overlap-and-add case for the signal model with SaS noise residual, it took approximately 38 min for each 131072 samples long superframe to be processed and 25 min for each 32768 samples long superframe, resulting in approximately 10 hours and 27 hours of processing time, respectively. When the song was processed as a whole it took around 11 hours. However, the greater memory requirements in the latter case compared to those of the overlap-and-add method make the latter method with 131072 samples long superframe a good compromise between speed and memory requirements. Not to mention that the overlap-and-add method can be exploited for parallel processing. The corresponding processing times for the signal model with Gaussian white noise residual are considerably smaller (i.e., 2 min for 131072 samples long superframes, 1 min for 32768 samples long superframes and 45 min for the full recording), since no additional effort is needed to estimate the SaS model parameters, and especially ρ_l.

6. CONCLUSIONS

A musical audio denoising technique has been proposed where the music signal is modeled by two MDCT bases in the frequency domain and the residual noise is modeled by means of an α-stable distribution. MCMC inference has been used to estimate all the parameters. The experimental results on musical excerpts from raw noisy recordings of Greek folk songs processed either as a whole or in superframes and overlap-and-add reconstruction demonstrate that the α-stable noise assumption is more suitable than the Gaussian white noise one. Moreover, the overlap-and-add method reduces memory requirements. The proposed method can be exploited to denoise old recordings maintained by cultural archives, music recording and publishing companies, or broadcasting corporations.

Acknowledgments

This research has been co-financed by the European Union

(European Social Fund - ESF) and Greek national funds through the Operation Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALIS - UOA - ERASITECHNIS MIS 375435.

7. REFERENCES

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

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

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

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

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

[6] 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.

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

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

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

[10] 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.

[11] C. Kereliuk and P. Depalle, "Sparse atomic modeling of audio: A review," in *Proc. 14th Int. Conf. Digital Audio Effects (DAFx-11)*, Paris, France, 2011.

[12] M. Plumbey, T. Blumensath, L. Daudet, R. Gribonval, and M. E. Davies, "Sparse representations in audio and music: From coding to source separation," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 995–1005, 2010.

[13] 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.

[14] K. Siedenburg and M. Dörfler, "Audio denoising by generalized time-frequency thresholding," in *Proc. 45th Int. Conf. AES*, Helsinki, Finland, 2012.

[15] D. Salas-Gonzalez, E. E. Kuruoğlou, 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.

[16] N. Bassiou, C. Kotropoulos, and E. Koliopoulou, "Symmetric α-stable sparse linear regression for musical audio denoising," in *Proc. 8th Int. Symposium Image and Signal Processing, and Analysis*, Trieste, Italy, September 4-6 2013, pp. 375–380.

[17] 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.

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

[19] 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.

[20] 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.

[21] 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.

[22] 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.

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

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