Greek Folk Music Denoising Under a Symmetric α-Stable Noise Assumption

Nikoletta Bassiou Constantine Kotropoulos Ioannis Pitas
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki 541 24, GREECE
Email: {nbassiou,costas,pitas}@aiia.csd.auth.gr

***Abstract*—The noise in musical audio recordings is assumed to obey an α-stable distribution. A sparse linear regression framework with structured priors is elaborated. Markov Chain Monte Carlo is used to infer the clean music signal model and the α-stable noise distribution parameters. 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. Experiments on noisy Greek folk music excerpts demonstrate better denoising under the α-stable noise assumption than the Gaussian white noise one, and when processing is performed in segments rather than in full recordings.**

I. Introduction

The Gaussian assumption is not suitable for modeling signal corruption which involves outliers, impulsiveness, and asymmetric characteristics [1], [2]. The α-stable distributions are more accurate models for 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, because their probability density function (PDF) cannot be analytically described in general. A particular mathematical representation was exploited to infer the α-stable parameters using the Gibbs sampler [5], while Monte Carlo Expectation-Maximization and Markov Chain Monte Carlo (MCMC) methods were introduced in [6], where the Scale Mixture of Normals (SMiN) representation of α-stable distributions was exploited. The SMiN property was also used to model symmetric α-stable (SaS) disturbances within a Gibbs Metropolis sampler [7]. More recently, a random walk MCMC approach for Bayesian inference in stable distributions was introduced resorting to a numerical approximation of the likelihood function [8]. An analytical approximation of the positive α-stable distribution by 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].

Frequently, the distortions in speech and music signals are localized, such as the CD scratches, when the signal is corrupted by impulsive noise (e.g., clicks) or when the waveform is truncated beyond a threshold (i.e., clipped) as well as when packet losses occur in cordless phones or voice over IP [11]. The distorted samples can be treated as missing and reconstruction algorithms could be employed to reconstruct the missing samples. Substantial efforts have been made to restore audio signals corrupted by clicks due to old recordings or scratched CDs by resorting to either autoregressive models [12], [13], Bayesian estimation of the corrupted samples [14], neural networks [15] or audio inpainting [11]. Non-negative matrix factorization was proposed as an alternative to Short Time Spectral Attenuation (STSA) for the digital curation of the musical heritage [16].

Here, the distortions in musical audio recordings is assumed to obey a SaS distribution model, extending the study [17] where a Gaussian white noise is assumed. The signal is modeled by two Modified Discrete Cosine Transform (MDCT) bases with the first basis describing the tonal parts of the signal and the second describing its transient parts [17]. Sparsity is enforced to the expansion coefficients of each MDCT base by means of binary indicator variables with structured priors as in [17]. A standard MCMC technique is employed to estimate the signal and the $\alpha-$stable noise parameters, following similar lines to [8], [18]. 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, extending the preliminary work [19]. The experimental results demonstrate a superior performance with respect to the power of the noise remaining after denoising and the acoustic perception of the denoised recordings, when the noise is assumed to be SaS and the musical recording is reconstructed using the overlap-and-add method.

The paper is organized as follows. In Section II, the signal modeling is presented, while in Section III the $\alpha-$stable model and the inference of $\alpha-$stable model parameters is elaborated. In Section IV, experimental results are discussed and conclusions are drawn in Section V.

II. 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 [17]. Tones and transients are captured by decomposing the audio signal into two types of MDCT atoms [20], while noise obeys a SaS distribution. Let $\mathbf{\Phi}_1 = [\Phi_{1,1}, \ldots, \Phi_{1,N}] \in \mathbb{R}^{N\times N}$ be the MDCT base with long frame length l_{frame_1} for representing the tonals and $\mathbf{\Phi}_2 = [\Phi_{2,1}, \ldots, \Phi_{2,N}] \in \mathbb{R}^{N\times N}$ be the MDCT base with short frame length l_{frame_2} for representing the transients. Obviously, $N = l_{frame_i} \times n_{frame_i}$, $i = 1, 2$. The atoms of either basis $\Phi_{i,k}$ 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, $\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2 \in \mathbb{R}^N$ be coefficient vectors, and $\mathbf{e} \in \mathbb{R}^{N\times 1}$ be a noise vector comprising independent identically distributed (i.i.d.) random variables (r.vs.) 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 audio signal model $\mathbf{x} \in \mathbb{R}^{N\times 1}$ is expressed as:

$$\mathbf{x} = \mathbf{\Phi}_1\tilde{\mathbf{s}}_1 + \mathbf{\Phi}_2\tilde{\mathbf{s}}_2 + \mathbf{e}. \quad (1)$$

The product property of the SaS distribution [3] suggests that the elements of $\mathbf{e}$, e_l, are equivalently represented by a Gaussian r.v. conditionally independent on the auxiliary positive stable r.v. ρ_l [18]:

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

The two vectors $\tilde{\mathbf{s}}_1 = [\tilde{s}_{1,1}, \ldots, \tilde{s}_{1,N}]^T$ and $\tilde{\mathbf{s}}_2 = [\tilde{s}_{2,1}, \ldots, \tilde{s}_{2,N}]^T$ are sparse, because 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 [17]. The parameters of the underlying clean signal model are estimated by means of MCMC inference methods. This means that appropriate conjugate priors should be 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 [17]:

$$p(\tilde{s}_{i,k}) = (1 - g_{i,k})\,\delta_0(\tilde{s}_{i,k}) + g_{i,k}\,\mathcal{N}(\tilde{s}_{i,k}|0, v_{i,k}) \quad (3)$$

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 $j = 1, 2, \ldots, l_{frame_i}$ by a Butterworth low-pass filter with filter order ν_i, cut-off frequency ω_i, and gain η_i:

$$h_{i,k} = \frac{\eta_i}{\left(1 + \frac{j-1}{\omega_i}\right)^{\nu_i}}. \quad (4)$$

2) Indicator variable priors: 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 [17]. 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 $\mathcal{B}(\pi_2|a_{\pi_2}, b_{\pi_2})$.

3) Gain parameter prior: The gain parameter η_i of the filter in (4) is given a Gamma conjugate prior, $p(\eta_i|a_{\eta_i}, b_{\eta_i}) = \mathcal{G}(\eta_i|a_{\eta_i}, b_{\eta_i})$ [17].

B. MCMC Inference

The parameters $\boldsymbol{\theta} = \{\tilde{\mathbf{s}}_i, v_i, \eta_i, P_{i,00}, P_{i,11}\}_{i=1,2} \cup \{\pi_2, \rho_i\gamma^2\}$ are sampled from their posterior distribution using the following MCMC scheme [17].

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 alternatively sampled one after the other. The likelihood of the observed audio signal $\mathbf{x}$ is written as follows

$$p(\mathbf{x}|\boldsymbol{\theta}) \sim \exp\left(-\frac{1}{2\gamma^2}\left\|\mathbf{\Sigma}_\rho(\mathbf{x} - \mathbf{\Phi}_1\tilde{\mathbf{s}}_1 - \mathbf{\Phi_2}\tilde{\mathbf{s}}_2)\right\|^2\right) \quad (5)$$

where $\mathbf{\Sigma}_\rho$ 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 $\tilde{\mathbf{x}}_{1|2} = \mathbf{\Phi}_1^\top(\mathbf{x} - \mathbf{\Phi}_2\tilde{\mathbf{s}}_2)$ or $\tilde{\mathbf{x}}_{2|1} = \mathbf{\Phi}_2^\top(\mathbf{x} - \mathbf{\Phi}_1\tilde{\mathbf{s}}_1)$, and $\tilde{\mathbf{e}}_i = \mathbf{\Phi}_i^\top\mathbf{e}$. 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, $g_{i,k}^{(l)}$ 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 $\tilde{s}_{i,k}^{(l)}$ is sampled from $p(\tilde{s}_{i,k}|g_{i,k}^{(l)}, v_i, \rho_i\gamma^2, \tilde{x}_{i|-i,k})$. A hypothesis testing problem is set to estimate the first posterior probability for $g_{i,k}$ [21]:

$$H_1 : g_{i,k} = 1 \Longleftrightarrow \tilde{x}_{i|-i,k} = \tilde{s}_{i,k} + \tilde{e}_{i,k} \quad (6)$$
$$H_0 : g_{i,k} = 0 \Longleftrightarrow \tilde{x}_{i|-i,k} = \tilde{e}_{i,k}. \quad (7)$$

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

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

where

$$\tau_{i,k} = \sqrt{\frac{\rho_i\gamma^2}{\rho_i\gamma^2 + v_{i,k}}}\exp\left(\frac{\tilde{x}_{i|-i,k}^2 v_{i,k}}{2\rho_i\gamma^2(\rho_i\gamma^2 + v_{i,k})}\right) \times\frac{p(g_{i,k} = 1|g_{i,-k}, \theta_{g_i})}{p(g_{i,k} = 0|g_{i,-k}, \theta_{g_i})}. \quad (8)$$

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

$$p(\tilde{s}_{i,k}|g_{i,k}, v_i, \rho_i\gamma^2, \tilde{x}_{i|-i,k}) = (1 - g_{i,k})\delta_0(\tilde{s}_{i,k}) + g_{i,k}\mathcal{N}(\tilde{s}_{i,k}|\mu_{\tilde{s}_{i,k}}, \sigma^2_{\tilde{s}_{i,k}}) \quad (9)$$

where

$$\sigma^2_{\tilde{s}_{i,k}} = \left(\frac{1}{\rho_i\gamma^2} + \frac{1}{v_{i,k}}\right)^{-1} \quad (10)$$
$$\mu_{\tilde{s}_{i,k}} = \left(\frac{\sigma_{\tilde{s}^2_{i,k}}}{\rho_i\gamma^2}\right)\tilde{x}_{i|-i,k}. \quad (11)$$

3) Updating of v_i *using Gibbs sampling:* The conditional posterior distribution of $v_{i,k}$ is given by $p(v_{i,k}|g_{i,k}, \tilde{s}_{i,k}, h_{i,k}) = (1 - g_{i,k})\ \mathcal{IG}(v_{i,k}|a_i, h_{i,k}) + g_{i,k}\,\mathcal{IG}\left(v_{i,k}\middle|\frac{1}{2} + a_i, \frac{\tilde{s}^2_{i,k}}{2} + h_{i,k}\right)$ [17].

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

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

5) Updating of η_i using Gibbs sampling: The full posterior distribution of the gain parameter η_i, which is given a Gamma conjugate prior, is $p(\eta_i|v_i) = \mathcal{G}\left(\eta_i\middle|Na_i+a_{\eta_i},\sum_k \frac{1}{1+\left(\frac{j-1}{\omega_i}\right)^{\nu_i}v_{i,k}}+b_{\eta_i}\right)$ [17].

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 (M-H) algorithm as described in [17] with the corresponding Beta distributions as proposed distributions.

III. $\alpha-$STABLE MODEL PARAMETER ESTIMATION

Similarly to the signal model, in order to estimate the unknown SaS parameters of the noise model (2), 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.

A. 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 is $\mathcal{N}\left(\frac{\frac{1}{\gamma^2}\sum_{l=1}^N\frac{e_l}{\rho_l}+\sigma_\delta m_\delta}{\frac{1}{\gamma^2}\sum_{l=1}^N\frac{1}{\rho_l}+\sigma_\delta},\frac{1}{\frac{1}{\gamma^2}\sum_{l=1}^N\frac{1}{\rho_l}+\sigma_\delta}\right)$ [18]. The full conditional for γ^2, that has an inverse Gamma conjugate prior [19], is the inverse Gamma distribution $\mathcal{IG}\left(a_0+\frac{N}{2},\frac{1}{2}\sum_{l=1}^N(e_l-\delta)^2+b_0\right)$ [18].

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.

(1) At each iteration t a candidate point α^{new} for α is generated from a proposal symmetric distribution $q(\cdot|\cdot)$. That is, $\alpha^{new}\sim q(\alpha^{new}|\alpha^{(t)})$.
(2) $\mathcal{U}$ is generated from a uniform $(0,1)$ distribution.
(3) If $\mathcal{U}\leq A(\alpha^{new}|\alpha^{(t)})$, 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 SaS probability density functions calculated arithmetically as in [3], [24][1].

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

3) Estimating auxiliary variable ρ_l with rejection sampling: Rejection sampling is used to sample from the posterior distribution $p(\rho)$

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

The likelihood forms a valid rejection function as it is bounded from above $p(e_l|\delta,\rho_l\gamma^2)\leq\frac{1}{\sqrt{2\pi}|e_l-\delta|}\exp\left(-\frac{1}{2}\right)$. Hence, the following rejection sampler can be used to draw samples from ρ_l [18]:

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 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.

IV. EXPERIMENTAL RESULTS

Two sets of experiments were conducted. In the first set, 4 noisy instrumental musical excerpts ($\simeq 48s$ long each) from Greek folk songs from the region of Western Macedonia were used, which were recorded in outdoor festivities. A clarinet and a drum are playing in all excerpts. The second set of experiments was conducted on a 48s long excerpt from a vocal song devoted to New Year's Carol that was sung by Mrs. Athina Korsavidou and recorded in 1930. This song is included in the collection "Songs of Pontos" released by the Melpo Merlier Music Folklore Archive [25]. Each excerpt was sampled at 44.1 kHz resulting in $T=2^{21}=2097152$ samples. The excerpts were also segmented in 17 and 67 segments of 131072 and 32768 samples each, respectively. In both cases, the segments were overlapping by 1024 samples. A sine-bell window was used for analysis and overlap-and-add reconstruction of the full denoised signals.

The proposed denoising algorithm, described in Sections II and III, was tested for restoring the excerpts as a whole as well as restoring the segments 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 Jeffrey's non-informative distribution. (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 II and III were run for 300 iterations with a burn-in period of 240 iterations. The estimate of the clean signal was constructed 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), which expresses the ratio of the original noisy signal power to the estimated noise power [19]:

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

TABLE I. OUTPUT NI VALUES OF THE PROPOSED ALGORITHM FOR SAS NOISE RESIDUAL AND THE ALGORITHM IN [17] FOR GAUSSIAN WHITE NOISE RESIDUAL APPLIED ON THE MUSICAL EXCERPTS PROCESSED AS A WHOLE (NO OA) AND IN SEGMENTS BY MEANS OF OVERLAP-AND-ADD RECONSTRUCTION (OA_1: 131072 SAMPLES LONG, AND OA_2: 32768 SAMPLES LONG).

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
6	Kalantarts Kali Chronia (New Year's Carol)	32.78	26.99	31.32	87.64	87.68	87.65

The smaller NI value, the higher noise power removal is attained and thus the better denoising performance is obtained. The output NI values measured for the algorithm developed in Section II, when α-stable noise residual is assumed in (1), are listed in Table I 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 [17] that resorts to Gaussian noise residuals, are included.

As can be seen in Table I, the assumption of a SaS noise residual in (1) and the consequent modifications due to this assumption in the framework proposed in [17] yields better denoising than the assumption of a Gaussian white-noise residual. For the SaS noise residual assumption, the denoising performance is considerably improved, when the musical excerpts are processed in segments and reconstructed by the overlap-and-add method. A negligible improvement was noticed when the musical excerpts are processed in segments, when a Gaussian white-noise residual is assumed. These conclusions are also verified by listening to the denoised musical excerpts[2]. When a Gaussian white noise residual is assumed, the processed audio recordings still contain a considerable amount of noise together with some new artifacts. When a SaS noise residual is assumed, the recordings are free from noise, but some cracks are inserted. In Fig. 1, 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 [17] that resorts to a Gaussian noise residual (b1-b2). By comparing Fig. 1(a1) and Fig. 1(b1), it is seen that the proposed variant for the tonal layer yields similar results with the original algorithm in [17]. However, the performance of the two algorithms differs significantly for the transient layer. Indeed, more artifacts are present, when a Gaussian noise residual is assumed (Fig. 1(b2)) than when a SaS stable noise residual is assumed (Fig. 1(a2)).

The MCMC inference for the SaS parameters is shown in Fig. 2, 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 (i.e., superframe) of the musical excerpt processed 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 $\sqrt{\rho}\gamma \simeq 2.4$ and $\sqrt{\rho}\gamma \simeq 2.6$ for the overlap-and-add with 131072 and 32768 samples, respectively.

[2]https://www.dropbox.com/sh/quorg0j4uohnevd/AAAN_0ubKbZY4IjjPfS9wlF3a

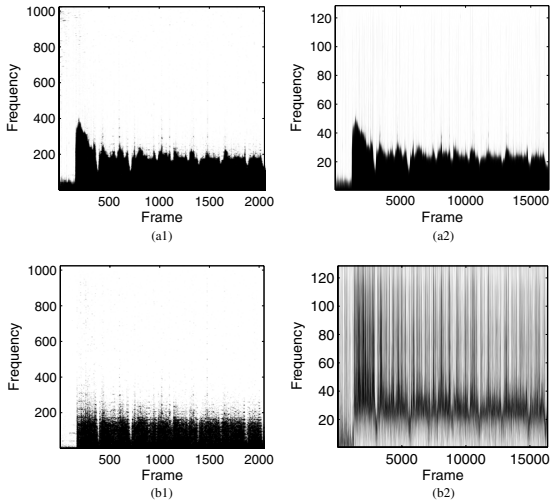


Fig. 1. Significance maps of the selected coefficients in $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ bases for the musical excerpt *4*. The maps show the MMSE estimates of the noise indicator variables g_1 and g_2 for: (a1)-(a2) SaS noise residual and (b1)-(b2) Gaussian white noise residual in (1). The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

The mean values for the stable parameter δ are of the order of 10^{-4} in all cases, as expected.

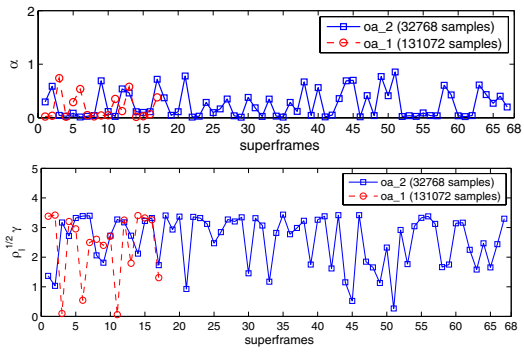
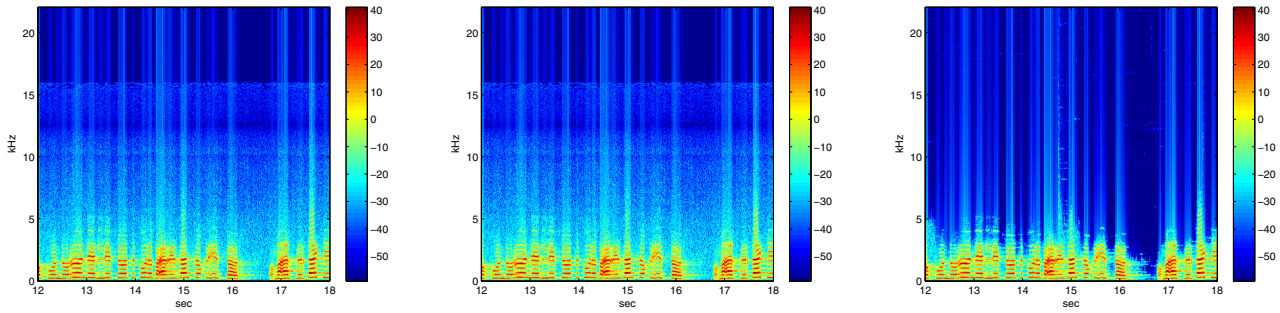


Fig. 2. Sampled values of the characteristic exponent α and standard deviation $\sqrt{\rho_l}\gamma$ of the SaS noise averaged across the iterations of the Gibbs sampler for each segment (i.e., superframe) in the overlap-and-add method.

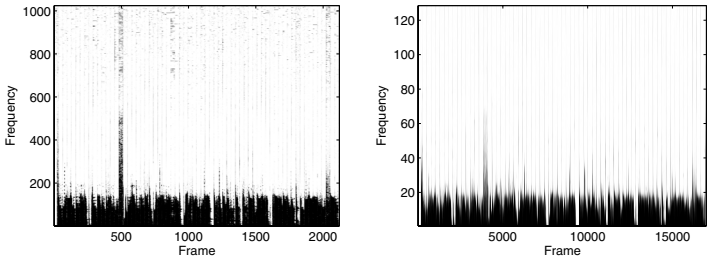
Spectrograms of a 6s long excerpt extracted from the 6th vocal song of Pontos that is devoted to New Year's Carol are shown in Fig. 3. In particular, the spectrogram of the original recording dating back to 1930 is shown in Fig. 3(a). The spectrograms of the denoised recordings that are reconstructed



(a) (b) (c)

Fig. 3. Spectrograms of a 6s long excerpt from the 6th 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.

by the overlap-and-add method, when either a Gaussian or a SaS noise residual is assumed are shown in Figs. 3(b) and (c). In the latter case, segments of 131072 samples were employed. The inspection of Fig. 3(c) reveals the superior denoising performance when a SaS noise residual is assumed.



(a) (b)

Fig. 4. Significance maps of the selected coefficients in $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ bases for the musical excerpt 6. The maps show the MMSE estimates of the noise indicator variables g_1 (a) and g_2 (b) for the SaS noise residual. The values range from 0 (white) to 1 (black).

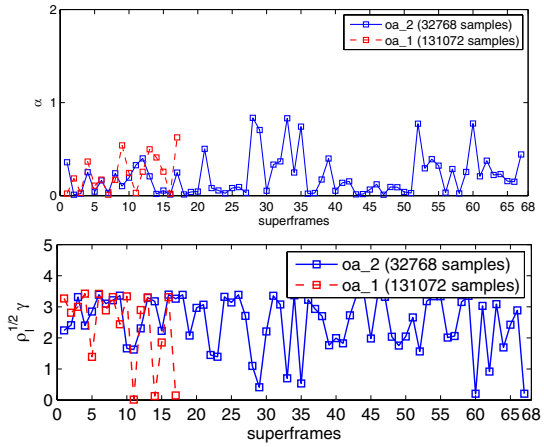


Fig. 5. Sampled values of the characteristic exponent α and standard deviation $\sqrt{\rho_l}\gamma$ of the SaS noise averaged across the iterations of the Gibbs sampler for each superframe (i.e., segment) in the overlap-and-add method.

The MMSE estimates of the indicator variables g_1 and g_2 for the vocal song are depicted in Figure 4 and they look alike those estimated for the instrumental song in Fig. 1(a1) and (a2).

The MCMC inference for 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 in Fig. 5 for each segment of the vocal song is reconstructed by the overlap-and-add method. The corresponding mean values are: $\alpha \simeq 0.23$ and $\alpha \simeq 0.21$ for the overlap-and-add with 131072 and 32768 samples, respectively. $\sqrt{\rho}\gamma \simeq 2.483$ and $\sqrt{\rho}\gamma \simeq 2.406$ for the overlap-and-add with 131072 and 32768 samples, respectively.

All the experiments were run on a Mac Core 2 Duo running at 2.4 GHz with 8 GB RAM. In the first set of experiments, it took approximately 38 min on average for each 131072 samples long segment to be processed in the overlap-and-add case for the signal model with the SaS noise residual. The processing time was reduced to 25 min for each 32768 samples long segment. It took approximately 10 and 27 hours to process each excerpt, respectively. When each instrumental recording was processed as a whole it took around 11 hours. In the second set of experiments, each 131072 samples long segment was processed in 43 min on average for the signal model with the SaS noise residual. For this model, the processing time was 28.35 min for each 32768 samples long segment. It took approximately 12 and 31 hours to conclude the processing of the vocal song. When the vocal song was processed as a whole it took around 17 hours. However, the greater memory requirements in the latter case than those of the overlap-and-add method make the overlap-and-add method with 131072 samples long segments a good compromise between speed and memory requirements. Not to mention that the overlap-and-add method is suitable for parallel processing. The processing times for the signal model with the Gaussian white noise residual are considerably smaller. For the instrumental recordings, 2 min for 131072 samples long segments, 1 min for 32768 samples long segments, and 45 min for the full recordings were required. For the vocal song the corresponding processing times were: 1.5 min for 131072 samples long segments, 30 s for 32768 samples long segments, and 25 min for the full recording. In this case, no additional effort is needed to estimate the SaS model parameters, and especially ρ_l.

V. Conclusions

A musical audio denoising technique has been proposed for music signals modeled by two MDCT bases in the frequency domain and residual noise modeled by 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 segments followed by an overlap-and-add reconstruction demonstrate that the α-stable noise assumption is more suitable than the Gaussian white noise one, though more computationally demanding. Moreover, the overlap-and-add reconstruction is found to reduce memory requirements and improve the performance with respect to the NI.

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

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

[4] J. P. Nolan, *Stable Distributions: Models for Heavy-Tailed Data*. Birkhaäuser, 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 α-stable distributions: A random walk MCMC approach," *Computational Statistics and Data Analysis*, vol. 51, no. 5, pp. 2688–2700, 2007.

[9] E. Kuruouğlu, "Analytical representation for positive α-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 α-stable time series models," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 2012, pp. 3737–3740.

[11] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. Plumbley, "Audio impainting," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 922–932, March 2012.

[12] A. Janssen, R. Veldhuis, and L. Vries, "Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 34, no. 2, pp. 317–330, April 1986.

[13] W. Etter, "Restoration of a discrete-time signal segment by interpolation based on the left-sided and right-sided autoregressive parameters," *IEEE Trans. Signal Processing*, vol. 44, no. 5, pp. 1124–1135, May 1996.

[14] S. J. Godsill and R. J. W. Rayner, *Digital Audio Restoration - A Statistical Model-Based Approach*. Berlin, Germany: Springer-Verlag, 1998.

[15] G. Cocchi and A. Uncini, "Subband neural networks prediction for on-line audio signal recovery," *IEEE Trans. Neural Networks*, vol. 13, no. 4, pp. 867–876, July 2002.

[16] G. Cabras, S. Canazza, P. L. Montessoro, and R. Rinaldo, "The restoration of single channel audio recordings based on non-negative matrix factorization and perceptual suppression rule," in *Proc. 13th Int. Conf. Digital Audio Effects*, Graz, Austria, September 6-10 2010, pp. 375–380.

[17] 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.

[18] 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.

[19] 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.

[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, no. 4, pp. 759–774, 1997.

[25] M. F. Dragoumis, *Songs of Pontos*, 2nd ed. Athens, Greece: Melpo Merlier Folklore Archive, 2006.