Restoration of Missing Data in Greek Folk Music by Interpolation Techniques

Paschalina Medentzidou and Constantine Kotropoulos
Department of Informatics, Aristotle University of Thessaloniki
Thessaloniki 54124, GREECE
Email: medentzidoup@iti.gr, costas@aiia.csd.auth.gr

***Abstract*—Music recordings often suffer from noise. The noisy segments may be treated as missing data. To restore them, one may employ interpolation techniques. A music signal is modeled as an autoregressive process, and three interpolation methods are developed that are based on maximum likelihood, Gibbs sampling, and Expectation Maximization. The aforementioned techniques are tested for restoration of missing data in vocal and instrumental Greek folk songs. Experimental results show that interpolation techniques based on maximum likelihood and Gibbs sampling offer better restoration results than Expectation Maximization.**

I. Introduction

Enhancement or restoration of audio recordings deals with problems broadly classified as localized and global disturbances, of which the most common are clicks (impulsive noise) and hiss (broadband noise), respectively [1]. Click suppression methods often use autoregressive (AR) models, whereas hiss reduction methods are frequently based, directly or indirectly, on some multi-rate approach. Music signals often are degraded by sudden, unexpected bursts of impulsive noise with random, but finite duration. Dirt, electrical interference or mechanical damage to the storage medium cause the loss of the original signal. The detection of the impulsive noise has been studied from many aspects [2], [3]. Once the impulsive noise is detected, the corrupted samples can be treated as missing.

Here, our interest is in applying statistical signal processing to the interpolation of a sequence of lost samples (i.e., a localized disturbance) that is based on the utilisation of a predictive and/or a probabilistic model of the audio signal and in particular to interpolation based on an AR model of the music signal employing a short-term prediction model [4]. 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 AR models [5], [6], Bayesian estimation of the corrupted samples [1], neural networks [7], or audio impainting [8].

In the following, we shall assume that the corrupted samples in a music signal have already been detected. Such samples will be restored by exploiting the information prior and after corruption, as shown in Figure 1. Let $\boldsymbol{y}_1 = \{x_i\colon 1 \leq i < m\}$ and $\boldsymbol{y}_2 = \{x_i : m+l \leq i \leq N\}$ denote the observed data and $\boldsymbol{z} = \{x_i : m \leq i < m+l\}$ be the missing data to be restored. The full recording can be represented as

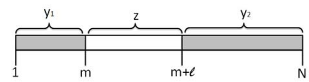

Fig. 1. The augmented data vector $\boldsymbol{x}$ consists of $\boldsymbol{y}_1, \boldsymbol{z}, \boldsymbol{y}_2$ where $\boldsymbol{z}$ is the vector of missing data having length l.

$\boldsymbol{x} = (\boldsymbol{y}_1^T |\, \boldsymbol{z}^T |\, \boldsymbol{y}_2^T)^T$. It is called the augmented data vector, hereafter. Three interpolation methods are developed, resorting to maximum likelihood, Gibbs sampling, and Expectation Maximization, respectively. The aforementioned techniques are tested for the restoration of missing data in instrumental and vocal Greek folk songs. It is demonstrated that maximum likelihood and Gibbs sampling offer a better restoration than Expectation Maximization. The interpolation problem can be stated as follows. Given the observed data $\boldsymbol{y}$, infer the value of the missing data $\boldsymbol{z}$.

A music signal is frequently modelled as an AR process of order p, i.e., for $p+1 \leq i \leq N$,

$$x_i = \sum_{j=1}^{p} \theta_j \, x_{i-j} + e_i \tag{1}$$

with the excitation sequence $\boldsymbol{e}$ made up of independent identically distributed (i.i.d.) Gaussian random variables of zero mean and standard deviation σ. Let $\mathbf{w} \in \mathbb{R}^{(N-p)\times 1}$ be the vector having elements x_i for $p+1 \leq i \leq N$. Then, the excitation sequence $\boldsymbol{e} \in \mathbb{R}^{(N-p)\times 1}$ can be obtained from

$$\boldsymbol{e} = \boldsymbol{w} - \boldsymbol{L}\,\boldsymbol{\theta} \tag{2}$$

where $\boldsymbol{\theta} \in \mathbb{R}^{p\times 1}$ is the vector of AR parameters and $\boldsymbol{L} \in \mathbb{R}^{(N-p)\times p}$ is defined as

$$\boldsymbol{L} = \begin{bmatrix} x_p & x_{p-1} & \cdots & x_2 & x_1 \\ x_{p+1} & x_p & \cdots & x_3 & x_2 \\ \vdots & & \cdots & & \vdots \\ x_{m-1} & x_{m-2} & \cdots & x_{m-p+1} & x_{m-p} \\ x_m & x_{m-1} & \cdots & x_{m-p+2} & x_{m-p+1} \\ \vdots & & \cdots & & \vdots \\ x_{N-1} & x_{N-2} & \cdots & x_{N-p+1} & x_{N-p} \end{bmatrix}. \tag{3}$$

Alternatively, the (full) excitation sequence $\boldsymbol{e} \in \mathbb{R}^{N\times 1}$ can be expressed in terms of the augmented data as

$$\boldsymbol{e} = \boldsymbol{K}\,\boldsymbol{x} \tag{4}$$

where $\boldsymbol{K} = \boldsymbol{K}(\boldsymbol{\theta}) \in \mathbb{R}^{N\times N}$ is the band diagonal Toeplitz matrix

$$\boldsymbol{K} = \begin{bmatrix} 1 & 0 & 0 & 0 & & 0 & 0 & 0 & 0 \\ -\theta_1 & 1 & 0 & 0 & & 0 & 0 & 0 & 0 \\ -\theta_2 & -\theta_1 & 1 & 0 & & 0 & 0 & 0 & 0 \\ & \ddots & \ddots & & & & & & \\ & & & & \ddots & \ddots & & & \\ 0 & 0 & 0 & 0 & & -\theta_1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & & -\theta_2 & -\theta_1 & 1 & 0 \\ 0 & 0 & 0 & 0 & & & -\theta_2 & -\theta_1 & 1 \end{bmatrix}. \tag{5}$$

The estimation of the model parameters (i.e., the missing data $\boldsymbol{z}$, the AR parameters $\boldsymbol{\theta}$, and the noise standard deviation σ) depends on the likelihood function of the excitation sequence given by

$$p(\boldsymbol{w}|\boldsymbol{\theta},\sigma) = p(\boldsymbol{e}) = \prod_{i=p+1}^{N} p(e_i) = (2\pi\sigma^2)^{-\frac{N-p}{2}} \exp\left(-\frac{\boldsymbol{e}^T\boldsymbol{e}}{2\sigma^2}\right). \tag{6}$$

Resorting to (2), the excitation energy is expressed in terms of the observed data $\boldsymbol{w}$ as:

$$\boldsymbol{e}^T\boldsymbol{e} = \boldsymbol{w}^T\boldsymbol{w} - 2\boldsymbol{w}^T\boldsymbol{L}\,\boldsymbol{\theta} + \boldsymbol{\theta}^T\boldsymbol{L}^T\boldsymbol{L}\,\boldsymbol{\theta}. \tag{7}$$

Let $\boldsymbol{K}^T\,\boldsymbol{K}$ be partitioned as [3]:

$$\boldsymbol{K}^T\,\boldsymbol{K} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{B}_1 & \boldsymbol{A}_{12} \\ \boldsymbol{B}_1^T & \boldsymbol{D} & \boldsymbol{B}_2^T \\ \boldsymbol{A}_{21} & \boldsymbol{B}_2 & \boldsymbol{A}_{22} \end{bmatrix}. \tag{8}$$

It can be shown that $\boldsymbol{K}^T\,\boldsymbol{K}$ is symmetric and band diagonal. In addition, it is Toeplitz after the p-th row [3]. Let also $\boldsymbol{A}$ and $\boldsymbol{B}$ be defined as

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{bmatrix}. \tag{9}$$

Let $\boldsymbol{y} = (\boldsymbol{y}_1^T|\boldsymbol{y}_2^T)^T$. An alternative expression for the excitation energy is obtained by employing the augmented data $\boldsymbol{x}$:

$$\boldsymbol{e}^T\boldsymbol{e} = (\boldsymbol{K}\,\boldsymbol{x})^T(\boldsymbol{K}\,\boldsymbol{x}) = \boldsymbol{y}^T\boldsymbol{A}\,\boldsymbol{y} + 2\boldsymbol{y}^T\boldsymbol{B}\,\boldsymbol{z} + \boldsymbol{z}^T\boldsymbol{D}\,\boldsymbol{z}. \tag{10}$$

II. Maximum Likelihood (ML) estimation

Let us elaborate the estimation of the model parameters by maximizing the likelihood (6) with respect to (wrt.) the AR parameters $\boldsymbol{\theta}$. This is equivalent to minimizing the excitation energy (7) wrt. $\boldsymbol{\theta}$, which yields the estimates:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{L}^T\,\boldsymbol{L})^{-1}\boldsymbol{L}^T\boldsymbol{w}. \tag{11}$$

Next, for fixed $\boldsymbol{\theta}$, minimizing (10) wrt. the missing data $\boldsymbol{z}$ yields the estimate

$$\hat{\boldsymbol{z}} = -\boldsymbol{D}^{-1}\,\boldsymbol{B}^T\,\boldsymbol{y} \tag{12}$$

where $\boldsymbol{D} \in \mathbb{R}^{l\times l}$, $\boldsymbol{B} \in \mathbb{R}^{(N-l)\times l}$, and $\boldsymbol{y} \in \mathbb{R}^{(N-l)\times 1}$. Finally, by differentiating (6) wrt. σ and equating with zero, we obtain the maximum likelihood estimate for scalar σ:

$$\hat{\sigma} = \sqrt{\frac{\boldsymbol{e}^T\,\boldsymbol{e}}{N}}. \tag{13}$$

The most computationally efficient procedure to derive $\boldsymbol{z}$ in (12) is to solve the system of equations $\boldsymbol{D}\boldsymbol{z} = -\boldsymbol{B}^T\boldsymbol{y}$, because $\boldsymbol{D}$ is band diagonal and symmetric with bandwidth equal to p [3]. Since the index of the starting missing sample m is greater than p, $\boldsymbol{D}$ is Toeplitz, a fact that yields computational savings whenever $2p^2 > l$. It is worth noting that $\boldsymbol{B}$ is highly sparse.

III. Gibbs sampler

Here, the Gibbs sampler samples vectors rather than scalars. The parameter space is partitioned into the missing data $\boldsymbol{z}$, the AR parameters $\boldsymbol{\theta}$, and the standard deviation σ. The Gibbs sampler draws random variates from the joint density $p(\boldsymbol{z},\boldsymbol{\theta},\sigma\,|\,\boldsymbol{y})$. Let $(\boldsymbol{z}^0,\boldsymbol{\theta}^0,\sigma^0)$ be a starting point in parameter space. If $\boldsymbol{z}^i \leftarrow p(\boldsymbol{z}\,|\,\boldsymbol{\theta}^{i-1},\sigma^{i-1},\boldsymbol{y})$ denotes a random sample $\boldsymbol{z}_i$ drawn from the conditional probability density function of $\boldsymbol{z}$ with all other parameters fixed at the i-th iteration, the sequence of variates

$$\begin{aligned} &\boldsymbol{z}^1 \leftarrow p(\boldsymbol{z}\,|\,\boldsymbol{\theta}^0,\sigma^0,\boldsymbol{y}) \\ &\boldsymbol{\theta}^1 \leftarrow p(\boldsymbol{\theta}\,|\,\sigma^0,\boldsymbol{z}^1,\boldsymbol{y}) \\ &\sigma^1 \leftarrow p(\sigma\,|\,\boldsymbol{z}^1,\boldsymbol{\theta}^1,\boldsymbol{y}) \\ &\boldsymbol{z}^2 \leftarrow p(\boldsymbol{z}\,|\,\boldsymbol{\theta}^1,\sigma^1,\boldsymbol{y}) \\ &\boldsymbol{\theta}^2 \leftarrow p(\boldsymbol{\theta}\,|\,\sigma^1,\boldsymbol{z}^2,\boldsymbol{y}) \\ &\vdots \\ &\sigma^i \leftarrow p(\sigma\,|\,\boldsymbol{z}^i,\boldsymbol{\theta}^i,\boldsymbol{y}) \end{aligned} \tag{14}$$

assumes a distribution in parameter space that asymptotically approaches the joint density $p(\boldsymbol{z},\boldsymbol{\theta},\sigma\,|\boldsymbol{y})$. In audio restoration, only the missing data component is of interest. The AR parameters and the standard deviation of the excitation sequence are ignored. Because the three variables are drawn from the joint density, one variable taken on its own is actually a sample from the marginal density, implying that the missing data are drawn from the predictive density $p(\boldsymbol{z}\,|\,\boldsymbol{y})$.

If we assume uniform prior probabilities on $\boldsymbol{z},\boldsymbol{\theta}$, and σ, then we can write:

$$p(\boldsymbol{z},\boldsymbol{\theta},\sigma\,|\,\boldsymbol{y}) = \frac{p(\boldsymbol{z},\boldsymbol{y}\,|\,\boldsymbol{\theta},\sigma)\;p(\boldsymbol{\theta},\sigma)}{p(\boldsymbol{y})} \propto p(\boldsymbol{z},\boldsymbol{y}\,|\,\boldsymbol{\theta},\sigma). \tag{15}$$

The conditional density of the missing data is then given by:

$$p(\boldsymbol{z}\,|\,\boldsymbol{\theta},\sigma,\boldsymbol{y}) = \frac{p(\boldsymbol{z},\boldsymbol{\theta},\sigma\,|\,\boldsymbol{y})}{\int_{\mathcal{Z}} p(\boldsymbol{z},\boldsymbol{\theta},\sigma\,|\,\boldsymbol{y})\,d\boldsymbol{z}} \tag{16}$$

where $\mathcal{Z} = \mathbb{R}^{l\times 1}$ is the domain of the missing data of length l. Similarly, the conditional density of the AR parameters is expressed as:

$$p(\boldsymbol{\theta}\,|\,\sigma,\boldsymbol{z},\boldsymbol{y}) = \frac{p(\boldsymbol{z},\boldsymbol{\theta},\sigma\,|\,\boldsymbol{y})}{\int_{\boldsymbol{\Theta}} p(\boldsymbol{z},\boldsymbol{\theta},\sigma\,|\,\boldsymbol{y})\,d\boldsymbol{\theta}} \tag{17}$$

where $\Theta = \mathbb{R}^{p\times 1}$, is the domain of the AR parameters. Finally, the conditional density of the standard deviation is given by:

$$p(\sigma \mid \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{y}) = \frac{p(\boldsymbol{z}, \boldsymbol{\theta}, \sigma \mid \boldsymbol{y})}{\int_0^\infty p(\boldsymbol{z}, \boldsymbol{\theta}, \sigma \mid \boldsymbol{y})\, d\boldsymbol{\sigma}}. \quad (18)$$

A. Conditional density of the missing data

By substituting (10) in (6), we obtain:

$$p(\boldsymbol{z}, \boldsymbol{\theta}, \sigma|\boldsymbol{y}) \propto \sigma^{-N} \exp\left[-\tfrac{1}{2\sigma^2}(\boldsymbol{y}^T \boldsymbol{A}\, \boldsymbol{y} + 2\boldsymbol{y}^T \boldsymbol{B}\, \boldsymbol{z} + \boldsymbol{z}^T \boldsymbol{D}\, \boldsymbol{z})\right]. \quad (19)$$

Integrating out the missing data in the denominator of (16) gives:

$$\int_{-\infty}^{\infty} p(\boldsymbol{z}, \boldsymbol{\theta}, \sigma \mid \boldsymbol{y})\, d\boldsymbol{z} \propto \sigma^{-N} \exp\left[-\tfrac{1}{2\sigma^2}\, \boldsymbol{y}^T \boldsymbol{A}\, \boldsymbol{y}\right]$$
$$\int_{-\infty}^{\infty} \exp\left[-\tfrac{1}{2\sigma^2}(2\boldsymbol{y}^T \boldsymbol{B}\, \boldsymbol{z} + \boldsymbol{z}^T \boldsymbol{D}\, \boldsymbol{z})\right]\, d\boldsymbol{z}. \quad (20)$$

Using the identity [9]:

$$\int_{-\infty}^{\infty} \exp\left(-\tfrac{1}{2}\, \boldsymbol{x}^T \boldsymbol{A}\, \boldsymbol{x} + \boldsymbol{J}^T \boldsymbol{x}\right)\, d\,x_1\, d\,x_2 ... d\,x_n$$
$$= \frac{(2\pi)^{n/2}}{|\boldsymbol{A}|^{1/2}} \exp\left(\tfrac{1}{2}\, \boldsymbol{J}^T \boldsymbol{A}^{-1} \boldsymbol{J}\right) \quad (21)$$

where $|\boldsymbol{A}|$ is the determinant of matrix $\boldsymbol{A}$ and $n = \dim(\boldsymbol{x})$ is the length of $\boldsymbol{x}$, (20) becomes

$$\int_{-\infty}^{\infty} p(\boldsymbol{z}, \boldsymbol{\theta}, \sigma \mid \boldsymbol{y})\, d\boldsymbol{z} \propto \sigma^{-(N-l)}$$
$$\exp\left[-\tfrac{1}{2\sigma^2}\, \left(\boldsymbol{y}^T \boldsymbol{A}\, \boldsymbol{y} - \boldsymbol{y}^T \boldsymbol{B}\, \boldsymbol{D}^{-1} \boldsymbol{B}^T \boldsymbol{y}\right)\right]. \quad (22)$$

The substitution of (19) and (22) into (16) yields:

$$p(\boldsymbol{z} \mid \boldsymbol{\theta}, \sigma, \boldsymbol{y}) = \sigma^{-l} \exp\left[-\tfrac{1}{2\sigma^2} Q_{\boldsymbol{z}}\right] \quad (23)$$

where

$$\begin{aligned} Q_{\boldsymbol{z}} &= \boldsymbol{z}^T \boldsymbol{D}\, \boldsymbol{z} + 2\boldsymbol{y}^T \boldsymbol{B}\, \boldsymbol{z} + \boldsymbol{y}^T \boldsymbol{B}\, \boldsymbol{D}^{-1} \boldsymbol{y} \\ &= (\boldsymbol{z} - \hat{\boldsymbol{z}})^T \boldsymbol{C}^{-1} (\boldsymbol{z} - \hat{\boldsymbol{z}}) \end{aligned} \quad (24)$$

with the inverse covariance matrix of the missing data given by

$$\boldsymbol{C}^{-1} = \frac{\boldsymbol{D}}{\sigma^2}. \quad (25)$$

The mode $\hat{\boldsymbol{z}}$ of the conditional density (23) is as in (12) i.e.,

$$\hat{\boldsymbol{z}} = -\boldsymbol{D}^{-1} \boldsymbol{B}^T \boldsymbol{y}. \quad (26)$$

To generate the missing samples, first the Cholesky decomposition of the band diagonal matrix $\boldsymbol{C}^{-1}$ is computed, i.e., $\boldsymbol{C}^{-1} = \boldsymbol{S}^T\, \boldsymbol{S}$. If $\boldsymbol{u} \in \mathbb{R}^{l\times 1}$ is a Gaussian random vector of zero mean and unit covariance matrix, we solve for $\overline{\boldsymbol{z}}$ such that $\boldsymbol{S}\, \overline{\boldsymbol{z}} = \boldsymbol{u}$ with a band LU decomposition. Finally, $\boldsymbol{z} = \hat{\boldsymbol{z}} + \overline{\boldsymbol{z}}$, where $\hat{\boldsymbol{z}}$ is given by (26).

B. Conditional density of the AR parameters

By substituting (7) in (6), we obtain:

$$p(\boldsymbol{z}, \boldsymbol{\theta}, \sigma|\boldsymbol{y}) \propto \sigma^{-N} \exp\left[-\tfrac{1}{2\sigma^2}(\boldsymbol{w}^T \boldsymbol{w} - 2\boldsymbol{w}^T \boldsymbol{L}\, \boldsymbol{\theta} + \boldsymbol{\theta}^T \boldsymbol{L}^T \boldsymbol{L}\, \boldsymbol{\theta})\right]. \quad (27)$$

Integrating out the AR parameters in the denominator of (17), we arrive at

$$\int_{-\infty}^{\infty} p(\boldsymbol{z}, \boldsymbol{\theta}, \sigma \mid \boldsymbol{y})\, d\boldsymbol{\theta} \propto \sigma^{-N+p}$$
$$\exp\left[-\tfrac{1}{2\sigma^2}\, \boldsymbol{w}^T \boldsymbol{w}\right] \exp\left[\boldsymbol{w}^T \boldsymbol{L}\, (\boldsymbol{L}^T \boldsymbol{L})^{-1} \boldsymbol{L}^T \boldsymbol{w}\right] \quad (28)$$

where $p = \dim(\boldsymbol{\theta})$. The substitution of (27) and (28) in (17) yields

$$p(\boldsymbol{\theta} \mid \sigma, \boldsymbol{z}, \boldsymbol{y}) \propto \sigma^{-p} \exp\left[-\tfrac{Q_{\boldsymbol{\theta}}}{2\sigma^2}\right] \quad (29)$$

where

$$\begin{aligned} Q_{\boldsymbol{\theta}} &= \boldsymbol{w}^T \boldsymbol{L}\, (\boldsymbol{L}^T \boldsymbol{L})^{-1} \boldsymbol{L}^T \boldsymbol{w} - 2\boldsymbol{w}^T \boldsymbol{L}\, \boldsymbol{\theta} + \boldsymbol{\theta}^T \boldsymbol{L}^T \boldsymbol{L}\, \boldsymbol{\theta} \\ &= (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \boldsymbol{C'}^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \end{aligned} \quad (30)$$

with the inverse covariance matrix of the reduced data given by

$$\boldsymbol{C'}^{-1} = \frac{\boldsymbol{L}^T \boldsymbol{L}}{\sigma^2}. \quad (31)$$

The mode $\hat{\boldsymbol{\theta}}$ of the conditional density (29) is as in (11), i.e.,

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{L}^T \boldsymbol{L})^{-1} \boldsymbol{L}^T \boldsymbol{w}. \quad (32)$$

To generate the AR parameters, a similar procedure to that of the generation of missing data is applied, i.e., $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} + \overline{\boldsymbol{\theta}}$, where $\hat{\boldsymbol{\theta}}$ is given by (32) and $\overline{\boldsymbol{\theta}}$ is such that $\boldsymbol{S}\, \overline{\boldsymbol{\theta}} = \boldsymbol{u}$ with $\boldsymbol{S}$ being the Cholesky factor of $\boldsymbol{C'}^{-1}$.

C. Conditional density of the standard deviation

Starting from

$$p(\boldsymbol{\sigma}, \boldsymbol{\theta}, \boldsymbol{z} \mid \boldsymbol{y}) = \sigma^{-N} \exp\left[-\frac{\boldsymbol{e}^T \boldsymbol{e}}{2\sigma^2}\right] \quad (33)$$

we can integrate out the standard deviation to arrive at:

$$\int_{-\infty}^{\infty} p(\boldsymbol{\sigma}, \boldsymbol{\theta}, \boldsymbol{z} \mid \boldsymbol{y})\, d\,\sigma \propto \left[\boldsymbol{e}^T \boldsymbol{e}\right]^{-\frac{N}{2}} \quad (34)$$

which is in the form of a Student-t distribution. It is seen that (34) does not depend on σ. Accordingly, by combining (33) and (34) we obtain the non-Gaussian density:

$$p(\sigma \mid \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{y}) \propto \left[\boldsymbol{e}^T \boldsymbol{e}\right]^{\frac{N}{2}} \sigma^{-N} \exp\left[-\frac{\boldsymbol{e}^T \boldsymbol{e}}{2\sigma^2}\right]. \quad (35)$$

To generate σ^i, start from a Gamma variate with $\frac{N-1}{2}$ degrees of freedom g^i, take the reciprocal of its square root, and scale the result, i.e., $\sigma^i = \sqrt{\frac{\boldsymbol{e}^T \boldsymbol{e}}{2}}\, \frac{1}{\sqrt{g^i}}$.

IV. Expectation Maximization- EM

Let us assume that $p(\boldsymbol{y})$ is constant, because $\boldsymbol{y}$ is fixed, and $p(\boldsymbol{\theta})$ is uniformly distributed to reflect our ignorance for the value of $\boldsymbol{\theta}$ in the absence of data. The starting point is to write out the predictive density:

$$p(\boldsymbol{z} \mid \boldsymbol{y}) \propto p(\boldsymbol{z},\, \boldsymbol{y}) = \frac{p(\boldsymbol{z},\, \boldsymbol{y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{\frac{p(\boldsymbol{z},\, \boldsymbol{y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\boldsymbol{z},\, \boldsymbol{y})}} \propto \frac{p(\boldsymbol{z},\, \boldsymbol{y} \mid \boldsymbol{\theta})}{p(\boldsymbol{\theta} \mid \boldsymbol{y},\, \boldsymbol{z})}, \quad (36)$$

where we have used $p(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{z}) = p(\boldsymbol{z}, \boldsymbol{y} \mid \boldsymbol{\theta}) \frac{p(\boldsymbol{\theta})}{p(\boldsymbol{z}, \boldsymbol{y})}$. Taking the logarithm of both sizes in (36), ignoring any additive constant, multiplying then both sides by $p(\boldsymbol{\theta}|\boldsymbol{y},\, \boldsymbol{z^*})$, and integrating with respect to $\boldsymbol{\theta}$, we arrive at

$$\log p(\boldsymbol{z}\mid\boldsymbol{y}) = \underbrace{\int_{\boldsymbol{\Theta}} \log p(\boldsymbol{z}\mid\boldsymbol{y},\boldsymbol{\theta})\, p(\boldsymbol{\theta}\mid\boldsymbol{y},\boldsymbol{z^*})\, d\boldsymbol{\theta}}_{Q(\boldsymbol{z},\boldsymbol{z}^*)} - \underbrace{\int_{\boldsymbol{\Theta}} \log p(\boldsymbol{\theta}\mid\boldsymbol{y},\boldsymbol{z})\, p(\boldsymbol{\theta}\mid\boldsymbol{y},\boldsymbol{z^*})\, d\boldsymbol{\theta}}_{H(\boldsymbol{z},\boldsymbol{z}^*)} \quad (37)$$

where $\boldsymbol{\Theta} = \mathbb{R}^p$ is the domain of the AR parameters. The left hand side of (37) is not a function of $\boldsymbol{\theta}$. Rao has shown that [10]:

$$H(\boldsymbol{z},\boldsymbol{z}^*) - H(\boldsymbol{z},\boldsymbol{z}) \le 0. \quad (38)$$

If $\boldsymbol{z}_i = \boldsymbol{z}^*$ is the estimation of missing data in the current iteration, then $\boldsymbol{z}_{i+1} = \boldsymbol{z}$ is chosen to maximize the function $Q(\boldsymbol{z},\boldsymbol{z}^*)$ in the next iteration. This value for $\boldsymbol{z}$ then becomes the new value for $\boldsymbol{z}^*$. Because of the inequality (38), each new value for $\boldsymbol{z}^*$ is guaranteed not to decrease the value of the predictive density $p(\boldsymbol{z}\mid\boldsymbol{y})$. The procedure is iterated until convergence, when the predictive density does not increase any more, at which point a supremum has been found. The Gaussian log likelihood can be written as:

$$\log p(\boldsymbol{z},\boldsymbol{y}\mid\boldsymbol{\theta}) = \frac{\boldsymbol{e}^T\boldsymbol{e}}{2\sigma^2} - \frac{N}{2}\log(2\pi\sigma^2). \quad (39)$$

From (2), we have $\boldsymbol{e} = \boldsymbol{w} - \boldsymbol{L}\,\boldsymbol{\theta}$. Let $\boldsymbol{L}_{i+1} = \boldsymbol{L}_{i+1}(\boldsymbol{y},\, \boldsymbol{z}_{i+1})$ be the matrix that results from missing data $\boldsymbol{z}_{i+1}$ and $\boldsymbol{L}_i = \boldsymbol{L}_{\,i}(\boldsymbol{y},\, \boldsymbol{z}_i)$ be the corresponding matrix that results when the missing data $\boldsymbol{z}_i$ are used. In the following, only functions of $\boldsymbol{\theta}$ and $\boldsymbol{z}$ are of concern.

A. Expectation

The expectation step involves integration over $\boldsymbol{\theta}$ in order to evaluate the function $Q(\boldsymbol{z},\boldsymbol{z}^*)$, i.e.,

$$Q(\boldsymbol{z_{i+1}},\boldsymbol{z_i}) = \int_{\boldsymbol{\Theta}} \log p(\boldsymbol{z}_{i+1},\boldsymbol{y}\mid\boldsymbol{\theta})\, p(\boldsymbol{z}_i,\boldsymbol{y}\mid\boldsymbol{\theta})\, d\boldsymbol{\theta} \quad (40)$$

where $p(\boldsymbol{z}_{i+1},\boldsymbol{y}\mid\boldsymbol{\theta}) = p(\boldsymbol{z}_{i+1}\mid\boldsymbol{y},\boldsymbol{\theta})\, p(\boldsymbol{y}\mid\boldsymbol{\theta})$ is used. The log probability in (40) is expressed as

$$\log p(\boldsymbol{z}_{i+1},\boldsymbol{y}\mid\boldsymbol{\theta}) \propto -\frac{(\boldsymbol{w}-\boldsymbol{L}_{i+1}\boldsymbol{\theta})^T(\boldsymbol{w}-\boldsymbol{L}_{i+1}\boldsymbol{\theta})}{2\sigma^2} - \frac{N}{2}\log(2\pi\sigma^2) \quad (41)$$

where $\boldsymbol{w} = (\boldsymbol{y}_1^T\mid\boldsymbol{z}_{i+1}^T\mid\boldsymbol{y}_2^T)^T$. The second conditional density in (40) is given by

$$p(\boldsymbol{z}_i,\boldsymbol{y}\mid\boldsymbol{\theta}) \propto (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left[-\frac{(\boldsymbol{v}-\boldsymbol{L}_i\,\boldsymbol{\theta})^T(\boldsymbol{v}-\boldsymbol{L}_i\,\boldsymbol{\theta})}{2\sigma^2}\right] \quad (42)$$

where $\boldsymbol{v} = (\boldsymbol{y}_1^T\mid\boldsymbol{z}_i^T\,|\boldsymbol{y}_2)^T$. Next, the integration of the product of (41) and (42) over the AR parameters $\boldsymbol{\theta}$ is outlined. For notation simplicity, $\boldsymbol{L_{i+1}}$ and $\boldsymbol{L_i}$ are denoted as $\boldsymbol{L}$ and $\boldsymbol{M}$, respectively. To arrive at a closed form expression for (40), the following substitutions are made:

$$(\boldsymbol{w}-\boldsymbol{L}\,\boldsymbol{\theta})^T(\boldsymbol{w}-\boldsymbol{L}\,\boldsymbol{\theta}) = (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^T\,\boldsymbol{L}^T\,\boldsymbol{L}\,(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) + \boldsymbol{w}^T\boldsymbol{w} - \boldsymbol{w}^T\boldsymbol{L}\,(\boldsymbol{L}^T\boldsymbol{L})^{-1}\boldsymbol{L}^T\boldsymbol{w} \quad (43)$$

$$(\boldsymbol{v}-\boldsymbol{M}\,\boldsymbol{\theta})^T(\boldsymbol{v}-\boldsymbol{M}\,\boldsymbol{\theta}) = (\boldsymbol{\theta}-\hat{\boldsymbol{\phi}})^T\,\boldsymbol{M}^T\,\boldsymbol{M}\,(\boldsymbol{\theta}-\hat{\boldsymbol{\phi}}) + \boldsymbol{v}^T\boldsymbol{v} - \boldsymbol{v}^T\boldsymbol{M}\,(\boldsymbol{M}^T\boldsymbol{M})^{-1}\boldsymbol{M}^T\boldsymbol{v} \quad (44)$$

where $\hat{\boldsymbol{\theta}} = (\boldsymbol{L}^T\boldsymbol{L})^{-1}\boldsymbol{L}^T\boldsymbol{w}$ and $\hat{\boldsymbol{\phi}} = (\boldsymbol{M}^T\boldsymbol{M})^{-1}\boldsymbol{M}^T\boldsymbol{v}$. Let $\boldsymbol{p} = \boldsymbol{\theta}-\hat{\boldsymbol{\phi}}$ and $\boldsymbol{q} = \boldsymbol{p}-(\hat{\boldsymbol{\theta}}-\hat{\boldsymbol{\phi}}) = \boldsymbol{\theta}-\hat{\boldsymbol{\theta}}$. Then, (40) takes the form

$$Q(\boldsymbol{z_{i+1}},\boldsymbol{z_i}) = \int_{\boldsymbol{\Theta}} A(\boldsymbol{p})\,\exp\left[-B(\boldsymbol{p})\right] d\boldsymbol{\theta} \quad (45)$$

where

$$A\,(\boldsymbol{p}) = -\frac{1}{2\sigma^2}\Big[\boldsymbol{p}^T\boldsymbol{L}^T\boldsymbol{L}\,\boldsymbol{p} - 2\,\boldsymbol{p}^T\boldsymbol{L}^T\boldsymbol{L}(\hat{\boldsymbol{\theta}}-\hat{\boldsymbol{\phi}}) + (\boldsymbol{w}-\boldsymbol{L}\,\hat{\boldsymbol{\phi}})^T(\boldsymbol{w}-\boldsymbol{L}\,\hat{\boldsymbol{\phi}})\Big] - \frac{N}{2}\log(2\pi\sigma^2) \quad (46)$$

$$B\,(\boldsymbol{p}) = -\frac{1}{2\sigma^2}\Big[\boldsymbol{p}^T\boldsymbol{M}^T\boldsymbol{M}\,\boldsymbol{p} + \boldsymbol{v}^T\boldsymbol{v} - \boldsymbol{v}^T\boldsymbol{M}\,(\boldsymbol{M}^T\boldsymbol{M})^{-1}\boldsymbol{M}^T\boldsymbol{v}\Big] - \frac{N}{2}\log(2\pi\sigma^2). \quad (47)$$

Using the identities [9]:

$$\int_{\mathbb{R}^q} \exp\left[-\boldsymbol{x}^T\boldsymbol{B}\,\boldsymbol{x}\right] dx = \frac{\pi^{\frac{q}{2}}}{\sqrt{\det\boldsymbol{B}}} \quad (48)$$

$$\int_{\mathbb{R}^q} \boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{y}\,\exp\left[-\boldsymbol{x}^T\boldsymbol{B}\,\boldsymbol{x}\right] dx = 0 \quad (49)$$

$$\int_{\mathbb{R}^q} \boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x}\,\exp\left[-\boldsymbol{x}^T\boldsymbol{B}\,\boldsymbol{x}\right] dx = \frac{\pi^{\frac{q}{2}}}{\sqrt{\det\boldsymbol{B}}}\frac{\mathrm{tr}(\boldsymbol{B}^{-1}\boldsymbol{A})}{2} \quad (50)$$

for $q = \dim(\boldsymbol{x})$ and assuming $\boldsymbol{y}$ independent of $\boldsymbol{x}$, we obtain:

$$Q(\boldsymbol{z_{i+1}},\boldsymbol{z_i}) = K(z_i)\left[\frac{\mathrm{tr}\big((\boldsymbol{M}^T\boldsymbol{M})^{-1}(\boldsymbol{L}^T\boldsymbol{L})\big)}{2} + \frac{\big(\boldsymbol{w}-\boldsymbol{L}\,\hat{\boldsymbol{\phi}}\big)^T\big(\boldsymbol{w}-\boldsymbol{L}\,\hat{\boldsymbol{\phi}}\big)}{2\sigma^2} - \frac{N}{2}\log(2\pi\sigma^2)\right] \quad (51)$$

where

$$K(z_i) = -\frac{(2\pi\sigma^2)^{-N/2}}{\sqrt{\det\,\boldsymbol{M}^T\boldsymbol{M}}}\exp\left[-\frac{(\boldsymbol{v}^T\boldsymbol{v}-\boldsymbol{v}^T\boldsymbol{M}\,(\boldsymbol{M}^T\boldsymbol{M})^{-1}\boldsymbol{M}^T\boldsymbol{v})}{2\sigma^2}\right]. \quad (52)$$

B. Maximization

Having found $Q(\boldsymbol{z}_{i+1},\boldsymbol{z}_i)$, its maximization wrt. $\boldsymbol{z}_{i+1}$ is sought by solving $\frac{\partial}{\partial\,\boldsymbol{z}_{i+1}}Q(\boldsymbol{z}_{i+1},\boldsymbol{z}_i) = 0$. This is equivalent to solving for $\boldsymbol{z}_{i+1}$ such that

$$\frac{1}{2}\mathrm{tr}\left[\left(\boldsymbol{L}_{\,i}^T\boldsymbol{L}_{\,i}\right)^{-1}\frac{\partial}{\partial\,\boldsymbol{z}_{i+1}}\left(\boldsymbol{L}_{\,i+1}^T\boldsymbol{L}_{\,i+1}\right)\right] \quad (53)$$

$$+\frac{1}{2\sigma^2}\frac{\partial}{\partial\,\boldsymbol{z}_{i+1}}(\boldsymbol{w}-\boldsymbol{L}_{i+1}\,\hat{\boldsymbol{\phi}})^T(\boldsymbol{w}-\boldsymbol{L}_{i+1}\,\hat{\boldsymbol{\phi}}) = 0 \quad (54)$$

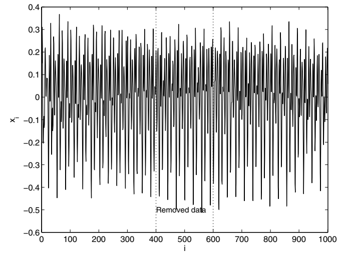

Fig. 2. 1000 samples from a Greek folk song from Pontus, where a woman sings a capella.

Similarly to (2) and (4), we claim $\boldsymbol{e} = \boldsymbol{w} - \boldsymbol{L}_{i+1}\,\hat{\boldsymbol{\phi}} = \boldsymbol{K}\,(\hat{\boldsymbol{\phi}})\,\boldsymbol{w}$. It can be showed that [3]:

$$\frac{1}{2\sigma^2}\frac{\partial}{\partial\,\boldsymbol{z}_{i+1}}(\boldsymbol{w} - \boldsymbol{L}_{i+1}\,\hat{\boldsymbol{\phi}})^T(\boldsymbol{w} - \boldsymbol{L}_{i+1}\,\hat{\boldsymbol{\phi}}) = \frac{1}{\sigma^2}(\boldsymbol{B}^T\boldsymbol{y} + \boldsymbol{D}\,\boldsymbol{z}_{i+1}). \tag{55}$$

The differentiation of the trace yields:

$$\frac{1}{2}\mathrm{tr}\left[\left(\boldsymbol{L}_i^T\boldsymbol{L}_{\,i}\right)^{-1}\frac{\partial}{\partial\,\boldsymbol{z}_{i+1}}\left(\boldsymbol{L}_{i+1}^T\boldsymbol{L}_{\,i+1}\right)\right] = \boldsymbol{T}\boldsymbol{z}_{i+1} + \boldsymbol{q} \tag{56}$$

where $\boldsymbol{T}$ is a symmetric band diagonal Toeplitz matrix with diagonal elements equal the sum of the corresponding diagonal of $(\boldsymbol{L}_i^T\boldsymbol{L}_i)^{-1}$. The term $\boldsymbol{q}$ depends on the reduced data and may be calculated efficiently by means of a convolution [3]. Accordingly, the M step of the EM algorithm solves the system of equations

$$\boldsymbol{T}\boldsymbol{z}_{i+1} + \boldsymbol{q} + \frac{1}{\sigma^2}(\boldsymbol{B}^T\boldsymbol{y} + \boldsymbol{D}\,\boldsymbol{z}_{i+1}) = \boldsymbol{0} \tag{57}$$

or equivalently the following band diagonal Toeplitz linear system of equations:

$$(\sigma^2\boldsymbol{T} + \boldsymbol{D})\boldsymbol{z}_{i+1} = -(\sigma^2\boldsymbol{q} + \boldsymbol{B}^T\boldsymbol{y}). \tag{58}$$

It is seen that the EM method solves (58) contrary to (26) that is solved by the ML. Both $\boldsymbol{T}$ and $\boldsymbol{q}$ can be obtained from $\sigma^2\left(\boldsymbol{L}^T\boldsymbol{L}\right)^{-1}$ [3].

V. Experiments

The three restoration techniques were applied to two Greek folk songs, namely a vocal song and an instrumental one[1]. The implementation of these techniques on real recordings presents certain difficulties, such as the need to cope with huge amounts of samples. For example, a song of 2 min duration sampled at a frequency $44.1\ kHz$ yields 5 million samples. In the following, we shall apply the restoration methods to music recording segments and not the full recordings. An important reason why we choose to process only parts of songs and not a whole song is the assumption made, that the part of the signal to be interpolated is stationary.

A segment of 1000 samples extracted from a Greek folk song is plotted in Fig. 2. This song is entitled $K\alpha\lambda\alpha\nu\tau\acute{\alpha}\rho\tau\varsigma\ \ \kappa\alpha\lambda\acute{\eta}\ \ \chi\rho o\nu\acute{\iota}\alpha$ (New Year's Carol). It was sung by Mrs. Athina Korsavidou and was recorded in 1930. The song is included in the collection "Songs of Pontos" released by the Melpo Merlier Music Folklore Archive [11]. Another segment of the same length from an instrumental Greek folk song is shown in Fig. 3. The song is entitled $K\alpha\lambda o\nu\upsilon\chi\tau\iota\acute{\alpha}$ (Good Night) and stems from the region of Western Macedonia. A clarinet and a drum is playing in it, which were recorded in outdoor festivities. Both signals in Figs. 2 and 3 are assumed AR processes of order $p = 40$.

[1] Code to reproduce the results can be found at: http://tinyurl.com/gt85tsq

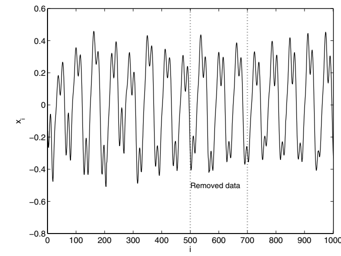

Fig. 3. 1000 samples of the instrumental Greek folk song under study.

From the first music segment shown in Fig. 2, 200 samples from the interval [400, 600] are removed. The restoration offered by the ML method, shown in Fig. 4, is the best with the restored signal being very close to the original signal. The restoration achieved by the Gibbs sampling after 600 repetitions is shown in Fig. 5. Note that the restored signal plotted is the mean of the last 50 iterations. The first 550 samples are considered as a burn-in period and are used in order to give the chain some time to start generating representative samples of the desired distribution. The restored signal looks like the original signal, but it has a smaller variability. The EM method yields the poor result shown in Fig. 6, which is attributed to the fact that the estimated excitation sequence in the gap approaches zero.

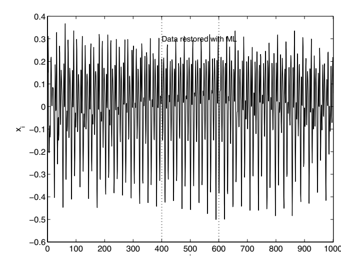

Fig. 4. Restoration of the song segment in Fig. 2 with ML.

From the second music segment shown in Fig. 3, 200 samples from the interval [500, 700] are removed. The ML method offers again the best restoration that is shown in Fig. 7. The restoration with Gibbs sampling is close to the original music segment, but not so accurate as can be seen in Fig. 8. The excitation sequence is vanished in the domain of missing samples with the EM method, yielding the interpolated signal plotted in Fig. 9.

The restoration quality index, i.e., the ratio of the restored signal energy over the original signal energy expressed in dB,

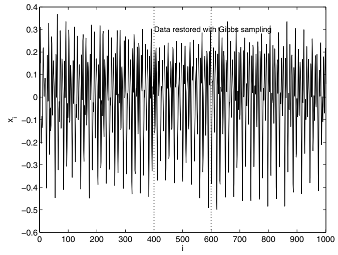

Fig. 5. Restoration of the song segment in Fig. 2 with Gibbs sampling.

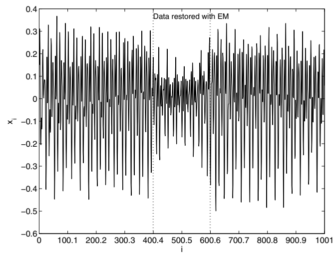

Fig. 6. Restoration of the song segment in Fig. 2 with EM.

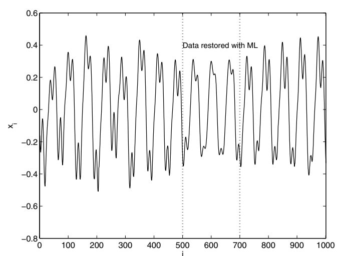

Fig. 7. Restoration of the song in Fig. 3 with ML.

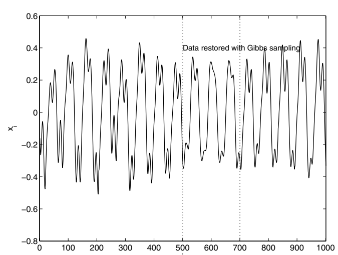

Fig. 8. Restoration of the song in Fig. 3 with Gibbs sampling.

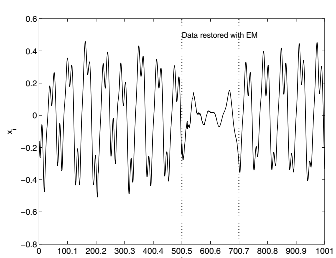

Fig. 9. Restoration of the song in Fig. 3 with EM.

has been used as a quantitative figure of merit to compare the three restoration techniques. Table I summarizes the measurements for both songs. The smallest absolute value indicates the best technique.

TABLE I
RESTORATION QUALITY INDEX OF THE VARIOUS RESTORATION METHODS.

	Method		
song	ML	Gibbs	EM
vocal	0.0199	0.3528	-2.2717
instrumental	-0.0146	0.0322	-7.9129

VI. CONCLUSIONS AND FUTURE WORK

In this paper, three restoration methods have been developed and tested for interpolating missing data in segments extracted from Greek folk songs. The ML estimation has shown to yield the best result. Future work includes the use of different assumptions for excitation sequences. Nothing precludes the application of the aforementioned techniques to the interpolation of missing data in image regions. An example is included in the code provided at github.

REFERENCES

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

[2] M. Wu and W. Fitzgerald, "Analytical approach to changepoint detection in laplacian noise," *IEE Proceedings Vision, Image, and Signal Processing*, vol. 142, no. 3, pp. 174–180, June 1995.

[3] J. Ó Ruanaidh and W. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing*, Springer, 1996.

[4] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, J. Wiley & Sons Ltd., Chichester, West Sussex, U.K., 4th edition, 2008.

[5] 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.

[6] 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.

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

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

[9] W. O. Straub, "A brief look at Gaussian integrals," Tech. Rep., Pasadena, California, 2009.

[10] R. Rao, *Linear Statistical Inference And Its Applications*, Wiley, New York, 1973.

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