Digital Signal Processing 50 (2016) 240–257

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Constrained non-negative matrix factorization for score-informed piano music restoration

F.J. Canadas-Quesada [a,*], P. Vera-Candeas [a], D. Martinez-Munoz [a], N. Ruiz-Reyes [a], J.J. Carabias-Orti [b], P. Cabanas-Molero [a]

[a] Telecommunication Engineering Department, University of Jaén, Alfonso X el Sabio, 28, 23700 Linares, Jaén, Spain
[b] Music Technology Group (MTG), Universitat Pompeu Fabra, Roc Boronat, 138, 08018 Barcelona, Spain

ARTICLE INFO

Article history:
Available online 25 January 2016

Keywords:
Audio restoration
Non-negative matrix factorization (NMF)
Sound source separation
Training
Single-activation
Score-based activation

ABSTRACT

In this work, we propose a constrained non-negative matrix factorization method for the audio restoration of piano music using information from the score. In the first stage (instrument training), spectral patterns for the target source (piano) are learned from a dataset of isolated piano notes. The model for the piano is constrained to be harmonic because, in this way, each pattern can define a single pitch. In the second stage (noise training), spectral patterns for the undesired source (noise) are learned from the most common types of vinyl noises. To obtain a representative model for the vinyl noise, a cross-correlation-based constraint that minimizes the cross-talk between different noise components is used. In the final stage (separation), we use the trained instrument and noise models in an NMF framework to extract the clean audio signal from undesired non-stationary noise. To improve the separation results, we propose a novel score-based constraint to avoid activations of notes or combinations that are not present in the original score. The proposed approach has been evaluated and compared with commercial audio restoration softwares, obtaining competitive results.

1. Introduction

Digital audio has more advantages with respect to physical degradation over time compared to analog audio. Currently, audio restoration for material degraded by vinyl noise (non-stationary noise) remains a challenging task, although it has been a widely investigated problem over the past decades [1–5]. It is well-known that vinyl noise is a major problem in music recordings because it produces annoying sounds due to mainly two causes, the material used in the manufacture of vinyl records, as imperfection in the pressing material and the physical wear. From a commercial perspective, audio restoration is an attractive tool because it has been applied in both the music industry (e.g., improving the quality of old recordings) [6–10] and forensic restoration (e.g., improving speech quality and intelligibility) [11–16].

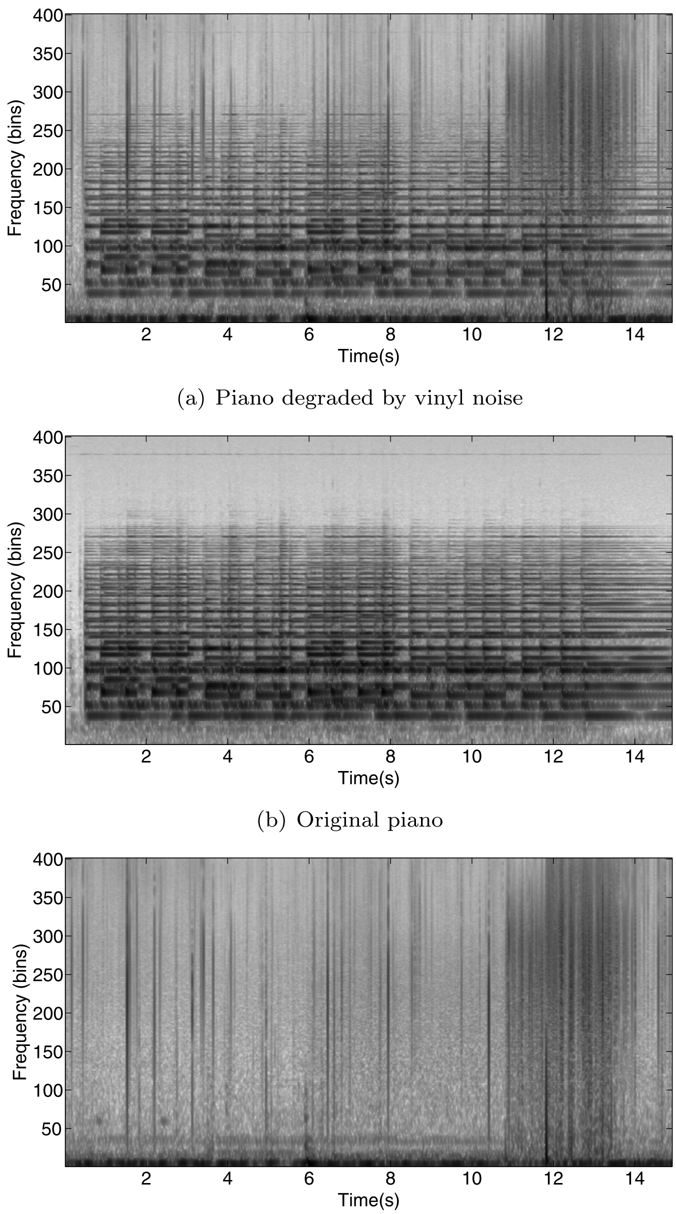
Audio restoration can be defined as the process of removing any degradation from the audio material to preserve the quality of the original material (see Fig. 1). The degradation can be classified as localized or global. Localized degradation is a discontinuity in the waveform which affects some time intervals of the audio, including clicks, crackles, pops, scratches, breakages, clipping, low frequency noise transients and buzz noise. This type of degradation is generally caused by dust, dirt, scratches or breakages on the surface of the record. In contrast, global degradation affects the entire waveform and it can be cited background noise (hiss), hum, rumble, wow and flutter and certain types of non-linear degradations such as speed variations or distortion [17].

Audio restoration approaches can be classified into two categories: frequency and time-domain methods. On the one hand, frequency-domain methods [1,18] are based on spectral subtraction schemes, in which the results are dependent on the estimated noise. On the other hand, time-domain methods [19,20] are based on the estimation of the statistical description of the audio events, and they incorporate signal models into the noise reduction.

Non-negative matrix factorization (NMF) is a more recent approach proposed for separating an acoustic source [21]. In fact, when the signal model is assumed to be non-negative, NMF provides a meaningful structure of the audio data, which in this case is obtained from the magnitude or power spectrograms. NMF is based on decomposing the spectrogram audio data into a sum of elementary spectral patterns (basis functions or components) with time-varying gains. In the basic form of NMF (baseline), unsupervised learning is used without applying constraints. This fact provides very little control over its behaviour because a local minima

* Corresponding author. Fax: +34953648508.
E-mail address: fcanadas@ujaen.es (F.J. Canadas-Quesada).

http://dx.doi.org/10.1016/j.dsp.2016.01.004
1051-2004/



(a) Piano degraded by vinyl noise

(b) Original piano

(c) Original vinyl noise

Fig. 1. Magnitude spectrogram of a 15-second piano excerpt degraded artificially (SNR = 0 dB) by vinyl noise (top), original piano (middle) from MAPS [38] and original vinyl noise (bottom). Fig. 1(c) shows typical degradations of vinyl noise, such as impulsive noise and background noise.

can be obtained from the decomposition using a set of bases that do not model spectral patterns typical of audio as those patterns can be found in the nature. Therefore, it appears natural to add explicit constraints to the factorization problem to retain certain semantics of the original signal, thereby providing meaningful and interpretable components. Thus, several constraints have been introduced to obtain NMF solutions that better fit certain expectancies. Among other proposed constraints, we can cite harmonicity [22–24], sparsity [25,26] or temporal continuity [25,27–29].

In this paper, we propose a constrained non-negative matrix factorization approach to separate the target source (piano) from the undesired source (vinyl noise). For this purpose, we propose to extend the cross-correlation penalty function developed by [28], which leads to two novel constraints: single-activation and score-based activation. The novelty of the first constraint lies in its use in the noise training stage to minimize the cross-talk between different noise spectral patterns, thereby reinforcing the representativity of each individual pattern. Thus, the cross-correlations between the components of the gain matrix are added as a regularization term to the global distortion. The main contribution of this paper is the score-based activation constraint because it uses information from the score (in MIDI format) to control the possible combinations of basis functions that can be active at one time. Concretely, concurrent activations of notes that do not occur in the score are penalized by a regularization term based on the cross-correlation between piano notes. Finally, the proposed constraints are included in an audio restoration framework with trained spectral patterns for piano sounds and vinyl noise. The results demonstrate that the use of the proposed constraints improves the audio quality in terms of the signal-to-distortion ratio (SDR), objective difference grades values (ODG) and MUSHRA scale.

The remainder of this paper is organized as follows. Section 2 reviews the background that is the basis for the proposed method. Section 3 details the proposed method. The evaluation results are presented in Section 4. Finally, we draw some conclusions and discuss future work in Section 5.

2. Background

2.1. Baseline NMF

Non-negative matrix factorization (NMF) [21] is a technique for multivariate data analysis that aims to obtain a parts-based representation of objects by imposing non-negative constraints. The problem addressed by NMF is as follows: given a matrix $\mathbf{X}$ of dimensions $F \times T$ with non-negative entries, it is possible to model the matrix as linear combinations of N elementary non-negative spectra (also called basis functions or components). Therefore, NMF is the problem of finding a factorization:

$$\mathbf{X} \approx \hat{\mathbf{X}} = \mathbf{BG} \tag{1}$$

where $\hat{\mathbf{X}}$ is the estimated matrix; $\mathbf{B} \in \mathbb{R}^{F \times N}$ is the matrix whose columns are the basis functions, which is also referred to as the basis matrix; and $\mathbf{G} \in \mathbb{R}^{N \times T}$ is a matrix of component gains for all frames. N is generally chosen such that $FN + NT \ll FT$, thereby reducing the dimensions of the data. In typical audio applications, the matrix $\mathbf{X}$ is chosen as a time-frequency representation (e.g., magnitude or power spectrogram), where $f = 1, \ldots, F$ denotes the frequency bin and $t = 1, \ldots, T$ is the time frame.

In the case of power and magnitude spectra, the parameters are restricted to be non-negative. Focusing on magnitude spectra, a common way to compute the factorization in eq. (1) is generally obtained by minimizing a cost function defined as

$$D(\mathbf{X}|\hat{\mathbf{X}}) = \sum_{f=1}^{F} \sum_{t=1}^{T} d(X_{f,t}|\hat{X}_{f,t}) \tag{2}$$

where $d(a|b)$ is a function of two scalar variables, and d is typically non-negative and takes a value of zero if and only if $a = b$. The most popular cost functions are the Euclidean distance $D_{\mathrm{EUC}}(\mathbf{X}|\hat{\mathbf{X}})$, the generalized Kullback–Leibler $D_{\mathrm{KL}}(\mathbf{X}|\hat{\mathbf{X}})$ and the Itakura–Saito $D_{\mathrm{IS}}(\mathbf{X}|\hat{\mathbf{X}})$ divergences. In this work, we used the generalized Kullback–Leibler divergence because it has been successfully applied in audio signal analyses [25,28,30].

An iterative algorithm based on multiplicative update rules is proposed in [21] to obtain the model parameters that minimize the cost function. Under these rules, $D_{\mathrm{KL}}(\mathbf{X}|\hat{\mathbf{X}})$ is non-increasing at each iteration, and the non-negativity of the bases and the gains is ensured. These multiplicative update rules are obtained by applying diagonal rescaling to the step size of the gradient descent algorithm. The multiplicative update rule for each scalar parameter Z