Audio restoration of solo guitar excerpts using a excitation-filter instrument model

J. Parras-Moral, F. J. Cañadas-Quesada, P. Vera-Candeas, N. Ruiz-Reyes
Telecommunication Engineering Department, University of Jaén, Spain
jpm00031@red.ujaen.es, fcanadas@ujaen.es, pvera@ujaen.es, nicolas@ujaen.es

ABSTRACT

This work proposes a denoising algorithm for musical instruments based on the use of an excitation-filter instrument model. Firstly, frequency patterns for the musical instrument are learned. These patterns are trained in advance from the RWC database and classified into harmonic and transient components. The harmonic patterns of the target instrument are modelled with an excitation-filter approach. Frequency patterns from the beginning of different notes (onsets) are also learned. Secondly, frequency patterns from noise are trained. Two different types of global degradations from vinyl audio (hum and hiss), apart from localized degradations from crackle noise, are used in this work. Two different types of global degradations from vinyl audio (hum and hiss), apart from localized degradations from click, crackle and scratch noise, are used in this work. Two databases (click+crackle+scratch+hiss and click+crackle+scratch+hiss+hum) are collected in order to obtain different subsets for training and testing. Finally, an NMF approach is applied to separate instrument signal and noise from noisy performances. The proposed approach is compared with some commercial algorithms when denoising a vinyl degraded guitar database. The separation measures indicate that the proposed approach obtains competitive results.

1. INTRODUCTION

The improvement of the quality for the audio material degraded by non-stationary noise in old recordings has been a widely investigated problem over the last years [1–4]. Nowadays, audio restoration is an attractive research field from a commercial viewpoint (e.g. albums or movies audio remastering) but it is still an unsolved problem because the quality of restored audio is quite dependent of algorithm parameters. Hence, it is necessary the judgment of subjects trained in audio to evaluate the quality of the audio processed.

Audio restoration is the process of removing any degradation to the audio material, which occurs as a result of the recording process, in order to preserve the quality of the original one. In general, any degradation can be classified into localized or global. A localized degradation, which affects only certain samples of the audio, can be described as impulsive noise such as click, crackle or scratch. It is usually caused by dust, dirt, scratches or breakages on the surface of the recording medium. A global degradation, which affects all samples of the audio, can be described as background noise. The main global degradations are known as hum and hiss noise. While hum noise models a $50-60Hz$ low frequency harmonic signal (caused by electrical noise), hiss noise models broadband noise (caused by ambient noise from the recording environment) [5].

Recent techniques, based on Non-negative Matrix Factorization (NMF) [6], has been successfully applied to a wide range of music analysis tasks [7–10]. Specifically, NMF is able to decompose a magnitude spectrogram as a product of two nonnegative matrices, $\mathbf{X} \approx \mathbf{B} \cdot \mathbf{G}$. Each column of the basis matrix $\mathbf{B}$ represents a spectral pattern from an active sound source. Each row of the gains matrix $\mathbf{G}$ represents the time-varying activations of a spectral pattern factorized in the basis matrix. In this paper, we propose a supervised NMF approach to restore the target audio by means of the removal or attenuation of any degradation in vinyl audio. This approach trains a set of spectral patterns that represent the target audio and the most common noise active in these recordings. The training audio of the target source is composed by samples of isolated notes from a spanish guitar instrument [11]. The spectral patterns from the guitar is trained both from the harmonic and onset components. The harmonic patterns are learned using a excitation-filter instrument model [10]. In the same way, the training audio of the noise is the concatenation of a wide set of public samples recorded from the most common types of vinyl noise [12–17]. Part of this material is not used in training to preserve the testing subset for vinyl noises. Some experiments have been developed in order to show the benefits of the use of instrument models and the trained spectral patterns of vinyl noises. Results are compared with some commercial approaches.

In this paper some proposals are shown. We propose the use of spectral patterns for the harmonic component of the instrument based on an excitation-filter model. The transient component of the instrument is taken into account training a set of spectral patterns from note onsets. Also, the vinyl noise spectral patterns are trained from some public samples. A model for the degraded audio composed by harmonic and transient components for the instrument and vinyl noise is developed. Separation is performed using an NMF algorithm to estimate the time-varying activations for spectral patterns from vinyl degraded signals. The instrument contribution of the mixed signal is obtained as a result of the separation process.

The paper is structured as follows: Section 2 reviews the state-of-the-art theory that is used in this paper; Section 3 shows the proposal of this work. The comparison of the obtained results with those obtained by other state-of-the-art methods are described at section 4 ; finally, we draw some conclusions and discuss future work in Section 5.

2. BACKGROUND

2.1 Augmented NMF parameter estimation

Standard Non-negative Matrix Factorization (NMF), developed by Lee and Seung [6], is a technique for multivariate data analysis where an input magnitude spectrogram, represented by matrix $\mathbf{X}$, is decomposed as a product of two non-negative matrices $\mathbf{B}$ and $\mathbf{G}$,

$$\mathbf{X} \approx \mathbf{B}\mathbf{G} \tag{1}$$

where $\mathbf{B}$ is the frequency basis and $\mathbf{G}$ represents the gains or activations of the active sources along the time, being $\hat{\mathbf{X}} = \mathbf{B}\mathbf{G}$ the approximation of the input matrix. The magnitude spectrogram $\mathbf{X}$, composed of T frames and F frequency bins, of a music signal consists of a set of time-frequency units $\mathbf{X}_{f,t}$ or $x(f,t)$.

Constraining parameters to be non-negative has been efficient in learning the spectrogram factorization models. In fact, this constraint has been widely used in SS [8, 18].

In the case of magnitude spectra, the parameters are restricted to be non-negative, then, a common way to compute the factorization is to minimize the reconstruction error between the observed spectrogram $x(f,t)$ and the modelled one $\hat{x}(f,t)$. This reconstruction error can be represented by a cost function.

The most used cost functions are the Euclidean (EUC) distance, the generalised Kullback-Leibner (KL) and the Itakura-Saito (IS) divergences. In this work, the KL cost function is used as is done in several systems [7, 9, 10].

An iterative algorithm based on multiplicative update rules is proposed in [6] to obtain the model parameters that minimize the cost function. Under these rules, $D_{KL}(x(f,t)|\hat{x}(f,t))$ is non-increasing at each iteration and it is ensured the non-negativity of the bases and the gains. These multiplicative update rules are obtained by applying diagonal rescaling to the step size of the gradient descent algorithm, more details can be found at [6]. The multiplicative update rule for each scalar parameter θ_l is given by expressing the partial derivatives of the $\nabla_{\theta_l} D_{KL}$ as the quotient of two positive terms $\nabla^{-}_{\theta_l} D_{KL}$ and $\nabla^{+}_{\theta_l} D_{KL}$:

$$\theta_l \leftarrow \theta_l \frac{\nabla^{-}_{\theta_l} D_{KL}(x(f,t)|\hat{x}(f,t))}{\nabla^{+}_{\theta_l} D_{KL}(x(f,t)|\hat{x}(f,t))} \tag{2}$$

The main advantage of the multiplicative update rule in eq. (2) is that non-negativity of the bases and the gains is ensured, resulting in an augmented non-negative matrix factorization (NMF) algorithm.

2.2 Multi-Excitation factorization Model (MEI)

The Multi-Excitation model proposed by Carabias et al. [10] is an extension of the source-filter model presented in [18]. This model achieves a good generalisation of the harmonic basis functions for a wide range of harmonic instruments [10], making its use a good alternative to obtain harmonic basis functions from a database of isolated sounds of the target instrument.

The source-filter model has origins in speech processing and sound synthesis. In speech processing, the excitation models the sound produced by the vocals cords, whereas the filter models the resonating effect of the vocal tract. In sound synthesis, excitation-filter (or source-filter) synthesis colors a spectrally rich excitation signal to get the desired sound.

The model proposed in [10] extend the source-filter model by defining the excitation as a weighted sum of instrument-dependent excitation patterns. Under this model, the spectrum of a note is generated by the harmonic excitation of the note multiplied by the filter transfer function of the instrument. Thus, the excitation $e_n(f)$ is different for each pitch and has harmonic nature. The pitch excitation is obtained as the weighted sum of excitation basis functions while the weights vary as the function of pitch.

Following this model, the pitch excitation can be obtained as

$$e_n(f) = \sum_{m=1}^{M} \sum_{i=1}^{I} w_{i,n} v_{i,m} G\left(f - m f_0(n)\right) \tag{3}$$

where $v_{i,m}$ is the i-th excitation basis vector (composed of M partials), and $w_{i,n}$ is the weight of the i-th excitation basis vector for pitch n. The basis functions $b_n(f)$ (or $\mathbf{B_{f,n}}$) are computed following the source-filter paradigm as

$$b_n(f) = h(f) e_n(f) \tag{4}$$

where $h(f)$ is the instrument-dependent instrument. Finally, the source-filter model with Multi-Excitation per Instrument (MEI) for magnitude spectra of the whole signal is the sum of instruments and pitches obtained as

$$\hat{x}(f,t) = \sum_{n} g_n(t) h(f) \sum_{m=1}^{M} \sum_{i=1}^{I} w_{i,n} v_{i,m} G\left(f - m f_0(n)\right) \tag{5}$$

where $n = 1, ..., N$ (N being the number of pitches), M represents the number of harmonics and I the number of considered excitations with $I << N$. Using a small number of excitation bases I reduces significantly the parameters of the model, which benefits the learning of parameters. The free parameter of the model are: the time gains $g_n(t)$ (or $\mathbf{G_{n,t}}$), the instrument filter $h(f)$, the basis excitation vectors $v_{i,m}$ and the excitation weigths $w_{i,n}$.

The framework presented in [6] can be used for MEI. For the sake of compact representation we present here the parameter update for the MEI model of (5). Multiplicative updates which minimize the divergence for each parameter of the MEI model are computed by substituting each parameter in eq. (2). More details can be obtained in [10].

3. DESCRIPTION

3.1 Signal factorization

Our proposal attempts to overcome the denoising problem learning in advance the harmonic and transient basis functions from the musical instrument and the spectral patterns from the vinyl noise. For that purpose, an objective function is defined to factorize a mixture spectrogram $X_{f,t}$ into three separated spectrograms, X_H (harmonic part of the musical instrument), X_T (transient part of the musical instrument) and X_N (vinyl noise part). We assume that each of them represents the specific spectral features demonstrated by the instrument and noise. In this manner, our factorization model is defined (see eq. 6),

$$\hat{\mathbf{X}} = \hat{\mathbf{X}}_H + \hat{\mathbf{X}}_T + \hat{\mathbf{X}}_N = \mathbf{B}_H\mathbf{G}_H + \mathbf{B}_T\mathbf{G}_T + \mathbf{B}_N\mathbf{G}_N \quad (6)$$

where all matrices are non-negative matrices.

In order to estimate basis functions or activation gains matrices, the iterative algorithm proposed in [6] can be applied. Using this algorithm, the update rule for the basis functions can be expressed as

$$\mathbf{B} = \mathbf{B} \odot \frac{\left[(\hat{\mathbf{X}})^{-1} \odot \mathbf{X}\right]\mathbf{H}'}{\mathbf{1}\mathbf{H}'} \quad (7)$$

where $'$ represents the transpose matrix operator, $\odot$ the element-wise multiplication of matrices, $\mathbf{1}$ is a all one elements matrix with F rows and T columns (or $\mathbf{1}_{f,t}$), $\mathbf{X}$ is the original spectrogram, $\hat{\mathbf{X}}$ is the modeled spectrogram and $\hat{\mathbf{X}}^{-1}$ is the inverse matrix regarding the modeled spectrogram. Eq. (7) can be used for each component of the proposed signal factorisation ($\mathbf{B}_H$, $\mathbf{B}_T$, $\mathbf{B}_N$).

The update rule for the activations gains can be written as

$$\mathbf{G} = \mathbf{G} \odot \frac{\mathbf{B}'\left[(\hat{\mathbf{X}})^{-1} \odot \mathbf{X}\right]}{\mathbf{B}'\mathbf{1}} \quad (8)$$

Both expressions are valid for each of the components represented in eq. (6).

In our approach, all basis functions ($\mathbf{B}_H$, $\mathbf{B}_T$, $\mathbf{B}_N$) are trained in advance from databases of guitar sounds or vinyl recorded noise.

3.2 Basis functions training

3.2.1 Instrument modeling for harmonic components

The model revised at section 2.2 requires to estimate the basis functions $b_n(f)$ for each note n defined in eq. (6) as the harmonic basis functions $\mathbf{B}_H$. The basis $b_n(f)$ are learned in advance by using the RWC database [11] as a training database of solo instruments playing isolated notes. Let the ground-truth transcription of the training data be represented by $r_n(t)$ as a binary time/frequency matrix. The frequency dimension represents the MIDI scale and time dimension t represents frames. $r_n(t)$ is known in advance for the training database, then it is used to initialize the gains for the training stage such that only the gain value associated with the active pitch n at frame t and played by instrument is set to unity, the rest of the gains are set to zero. Gains initialised to zero remain at zero because of the multiplicative update rules, and therefore the frame is represented only with the correct pitch.

The training procedure is summarised in Algorithm 1.

Algorithm 1 Training Harmonic Spectral Patterns

1. Compute $x(t, f)$ from a solo performance of the target instrument in the training database.
2. Initialise gains $g_n(t)$ with the ground truth transcription $r_n(t)$ and the rest of parameters $h(f)$, $v_{i,m}$ and $w_{i,n}$ with random positive values.
3. Update source-filter $h(f)$.
4. Update excitation basis vectors $v_{i,m}$.
5. Update the weights of the excitation basis vectors $w_{i,n}$.
6. Update gains $g_n(t)$.
7. Repeat steps 3-6 until the algorithm converges (or the maximum number of iterations is reached).
8. Compute basis functions $b_n(f)$ for the musical instrument from eq. (3) and (4) .

Basis function $b_n(f)$ are computed by this training algorithm resulting in a basis function for the complete pitch range n played by the instrument. The instrument-dependent basis functions $b_n(f)$ (or $\mathbf{B}_H$) are known and held fixed during the factorization process, and therefore, the factorization of new signals of the same instrument can be reduced to estimate the gains $g_n(t)$.

3.2.2 Learning transient basis functions

The transient spectral patterns from a musical instrument does not follow a harmonic behaviour. Here, our approach is to learn a representative set of transient basis functions from the note onsets of a training database. Again, the basis $\mathbf{B}_T$ are learned in advance by using the RWC database [11]. In order to initialize the gains for the training stage, lets define $ro(t)$ as a binary time/frequency vector that represents the frames in which a note onset is active. To obtain this vector the database of solo instruments playing isolated notes is annotated supposing that the transient components are active T_O frames from the beginning of each note. In our experiments, a value of $T_O = 5$ frames is used.

The training procedure is summarised in Algorithm 2, the number of transient basis functions is defined as O.

Algorithm 2 Training Transient Spectral Patterns

1. Compute $x(t, f)$ from a solo performance of the target instrument in the training database.
2. Initialise all gains $\mathbf{G}_T$ with random positive values for those frames in which a note onset is active using $ro(t)$.
3. Initialise transient basis functions $\mathbf{B}_T$ with random positive values.
4. Update basis functions $\mathbf{B}_T$.
5. Update gains $\mathbf{G}_T$.
6. Repeat steps 4-5 until the algorithm converges (or the maximum number of iterations is reached).

As in the harmonic case, transient basis functions $\mathbf{B}_T$ are known and held fixed during the factorization process.

3.2.3 Training basis functions from recorded vinyl noise

The vinyl noise used to train vinyl noise basis functions $\mathbf{B}_N$ was obtained from the concatenation of a wide set of public samples recorded from the most common types of vinyl noise [16] [17] [18] [19] [20] [21]. From this concatenation noise signal, two third of the total one was considered for training and the remainder for evaluation. Two groups of different degradations from vinyl noise are trained:

- clicks+crackles+scratches+hiss.
- clicks+crackles+scratches+hiss+hum.

The training procedure is summarised in Algorithm 3, the number of transient basis functions is defined as R.

Algorithm 3 Training vinyl Noise Spectral Patterns

1 Compute $\mathbf{X}$ from the training subset of the noise database.
2 Initialise all gains $\mathbf{G}_N$ with random positive values.
3 Initialise noise basis functions $\mathbf{B}_N$ with random positive values.
4 Update basis functions $\mathbf{B}_N$.
5 Update gains $\mathbf{G}_N$.
6 Repeat steps 4-5 until the algorithm converges (or the maximum number of iterations is reached).

Again, the two groups of noise basis functions $\mathbf{B}_N$ are known and held fixed during the factorization process.

3.3 Denoising application

In order to synthesize the denoised instrument signal, the magnitude instrumental spectrogram $\hat{\mathbf{X}}_H + \hat{\mathbf{X}}_T$ are estimated as the product of the factorization $\mathbf{B}_H\mathbf{G}_H + \mathbf{B}_T\mathbf{G}_T$. To assure a conservative reconstruction process, an instrumental mask $\mathbf{M}_J$ has been generated by means of Wiener filtering (the mask values are defined from 0 to 1).

Firstly, the magnitude spectrograms for the harmonic $\hat{\mathbf{X}}_H$ and transient $\hat{\mathbf{X}}_T$ components of the instrument are estimated using the factorization scheme proposed in eq. (6). In algorithmic approximation, the estimation of the instrumental spectrogram is detailed in Algorithm 4.

Algorithm 4 Estimation of instrumental components

1 Compute the magnitude spectrogram $\mathbf{X}$ of the degraded signal.
2 Initialise $\mathbf{G}_H$, $\mathbf{G}_T$ and $\mathbf{G}_N$ with random nonnegative values.
3 Initialise $\mathbf{B}_H$, $\mathbf{B}_T$ and $\mathbf{B}_N$ from the training algorithms.
4 Update $\mathbf{G}_H$.
5 Update $\mathbf{G}_T$.
6 Update $\mathbf{G}_N$.
7 Repeat steps 4-6 until the algorithm converges (or the maximum number of iterations is reached).
8 Compute the estimated instrumental spectrogram as $\hat{\mathbf{X}}_H + \hat{\mathbf{X}}_T$.

The instrumental mask is therefore defined as

$$\mathbf{M}_J = \frac{\hat{\mathbf{X}}_H + \hat{\mathbf{X}}_T}{\hat{\mathbf{X}}_H + \hat{\mathbf{X}}_T + \hat{\mathbf{X}}_N} \tag{9}$$

The phase information related to the instrumental signal is computed by multiplying the mask $\mathbf{M}_J$ with the complex spectrogram related to the degraded signal $x_J(t) + x_N(t)$. The inverse transform is then applied to obtain an estimation of the instrumental signal $\hat{x}_J(t)$.

4. EVALUATION

4.1 Material

Two test databases D1 and D2 of vinyl degraded guitar sounds were used to evaluate the performance of the proposal. Each database is composed of five degraded files. Each file [19–21] (see Table 1), 30-seconds duration, is created from a real-world Spanish guitar excerpt (with CD quality) degraded by typical noise in vinyl recordings. In the first database D1, degradations include clicks, crackles, scratches and hiss noise. In the second database D2, degradations include clicks, crackles, scratches, hiss and hum noise.

Identifier	Name
F1	Danza de los vecinos
F2	Iberia
F3	Albaicin
F4	Fuente y Caydal
F5	Rumba improvisada

Table 1. Real-world CD quality Spanish guitar excerpts used in experiments [19–21].

The degradation of the audio guitar excerpts was made using the concatenation signal of a wide set of public samples recorded from the most common types of vinyl noise [16] [17] [18] [19] [20] [21]. From this concatenation of vinyl noise, two thirds of the total was considered for training and the remainder for evaluation. So, different noise material was used for training and testing in order to validate the results. Specifically, the training material has durations of 228 seconds for clicks, crackles, scratches and hiss noise and 89 seconds for clicks, crackles, scratches, hiss and hum noise.

To evaluate different acoustic scenarios, the mixing process between guitar excerpts and vinyl noise was produced at 0, 5 and 10 dB of signal-to-noise ratio (see Table 2).

Name	Database	SNR (dB)
D1_0	D1	0
D1_5	D1	5
D1_10	D1	10
D2_0	D2	0
D2_5	D2	5
D2_10	D2	10

Table 2. Acoustic scenarios in the evaluation process.

4.2 Commercial audio restoration products

Three current and well-known commercial audio restoration products have been used to evaluate the performance of our proposal:

- Adobe Audition CS5.5 v4.0.

- Izotope RX 2 (Declicker, Decrackle, Denoiser and Hum removal).
- Waves V8 (X-Click, X-Crackle, X-Hum and Z-Noise).

Both Waves and Izotope plugins were used in Wavelab 6 audio editing and mastering suite from Steinberg [22]. Each audio restoration product has been manually tuned to provide the best results according to noise reduction and quality of the target audio.

4.3 Experimental setup

The proposed method has been evaluated by using the following parameters: frame size of $64ms$, hop size of $32ms$, frequency sampling rate of $44100Hz$, 100 iterations for NMF algorithm, number of transient basis functions $O = 10$ and number of vinyl noise basis functions $R = \{10, 100\}$ (see the following section). Sound source separation applications based on NMF algorithms usually adopt logarithmic frequency discretization. For example, uniformly spaced subbands on the Equivalent Rectangular Bandwidth (ERB) scale are assumed in [23]. In our method, we use the resolution of a quarter semitone by directly integrating the bins of the STFT similary to [10].

4.4 Results

For an objective evaluation of the performance of the separation method we use the metrics implemented in [23]. These metrics are commonly accepted by the specialised scientific community, and therefore facilitate a fair evaluation of the method. The metrics for each separated signal are the *Source to Distortion Ratio* (SDR), the *Source to Interference Ratio* (SIR), and the *Source to Artifacts Ratio* (SAR).

In an NMF framework, the unknown parameters are initialized randomly. Therefore, the spectra resulting from separation are different at each execution, giving different metric results per execution. Thus, the proposed method has been performed 50 times per audio file to demonstrate the statistical significance of the metrics. The 95% confidence interval for the metrics was always smaller than $1.1dB$ in the proposed method.

The SDR results for the denoised guitar signals when using the D1 and D2 databases at different SNRs are given in Table 3. The proposed methods are: P10 proposed method with $R = 10$ noise basis functions, UP10 unrealistic proposed method with $R = 10$ noise basis functions (the noise is directly trained from the same noise added to the degraded signal which is an unrealistic situation), P100 proposed method with $R = 100$ noise basis functions and UP100 unrealistic proposed method with $R = 100$ noise basis functions. The unrealistic approaches are used for estimating the loss produced in separation performance when training the vinyl noise in an implementation different from the real noise. The SDR value of the original input signal is also presented. As can be seen, Waves software obtains the best separation measures from the commercial restoration products. In our approach, the use of $R = 10$ bases is better than using $R = 100$, so we can conclude that the spectral richness of the vinyl noise can be captured with a reduced number of basis functions. Also, the proposed methods achieve better performance for the D2 database mainly because the hum noise is the most stable in frequency. Finally, we can state that our approach is competitive in relation to the commercial audio restoration software.

Name	Input	Audition	Izotope	Waves	P10	UP10	P100	UP100
D1_0	3.2	7.5	5.1	8.6	9.0	9.6	8.4	9.2
D1_5	8.3	11.8	11.2	11.7	12.4	12.9	11.4	12.2
D1_10	13.1	16.2	13.3	16.5	14.6	15.1	13.1	14.1
D2_0	4.7	-2.2	3.0	6.5	11.2	11.8	9.9	10.5
D2_5	9.7	-2.0	5.1	7.7	13.9	14.4	12.4	13.0
D2_10	14.6	-1.9	5.6	8.5	15.8	16.3	13.9	14.6

Table 3. *Denoised guitar SDR results in dB for D1 and D2 databases.*

The SIR results for the denoised guitar signals when using the D1 and D2 databases at different SNRs are given in Table 4. These results inform about the amount of noise present in the cleaned guitar. In all cases, the denoised signals with the proposed methods have less interferences from the vinyl noise.

Name	Input	Audition	Izotope	Waves	P10	UP10	P100	UP100
D1_0	3.3	8.7	8.7	11.7	11.5	12.3	11.1	12.3
D1_5	8.5	13.3	15.2	14.2	16.3	17.0	16.1	17.0
D1_10	13.3	18.3	20.6	20.0	20.6	21.1	20.4	21.2
D2_0	9.7	9.7	12.2	20.8	21.4	21.5	20.5	21.2
D2_5	14.7	14.3	17.6	21.8	25.4	25.4	24.8	25.5
D2_10	19.7	19.0	22.6	28.1	29.2	29.4	28.8	29.5

Table 4. *Denoised guitar SIR results in dB for D1 and D2 databases.*

The SIR results for the estimated vinyl noise component when using the D1 and D2 databases at different SNRs are given in Table 5. Now, the amount of original guitar eliminated from the denoised guitar is shown. On the contrary, in this case Audition and Waves approaches obtain much better results than the proposed approach for the D1 database.

Name	Audition	Izotope	Waves	P10	UP10	P100	UP100
D1_0	18.1	1.7	17.2	10.6	11.4	8.3	10.1
D1_5	19.8	6.7	23.4	5.5	6.4	3.1	4.9
D1_10	16.8	1.8	18.6	0.6	1.7	-1.9	-0.1
D1_0	-11.6	-8.5	-1.8	3.2	3.7	0.6	2.4
D2_5	-16.0	-10.7	-7.0	-1.7	-1.1	-4.1	-2.4
D2_10	-20.2	-14.8	-11.9	-6.1	-5.5	-8.4	-6.7

Table 5. *Estimated vinyl noise SIR results in dB for D1 and D2 databases.*

In order to give the reader the opportunity of listening the material a webpage for the results has been created. On this page, some audio examples (mixed, separated guitar and separated noise) from database D1 and D2 can be heard by the reader. The web page can be found at *http://dl.dropbox.com/u/22448214/SMC%202013/index.html*

5. CONCLUSIONS AND FUTURE WORK

In this work, a denoising technique based on an excitation-filter model for harmonic instruments is proposed. The instrumental part of the degraded signal is divided into harmonic and transient components and trained from the RWC database. The vinyl noise is trained from public recordings. Basis functions are fixed from the training algorithms and in the separation process the activation gains for each component are estimated following an NMF framework. The results show that the proposed approach are competitive in comparison with some commercial audio restoration softwares.

The main problem of the proposed approach is the similarity of the transient basis functions for the instrument and the spectral patterns of the localized degradations such as click, crackle and scratch noise. In our opinion, this issue causes the presence of instrument interferences in the estimated noise and, consequently, the loss of instrument signal in the denoised instrumental audio. This problem also occurs when training the vinyl noise from the original noise (UP10 and UP100 approaches).

For future work, an interesting idea to solve the interference problems can be the definition of sparseness and smoothness constraints [18] in the basis functions and activations gains of the signal factorization.

6. REFERENCES

[1] S. Godsill, P. Wolfe, and W. Fong, "Statistical model-based approaches to audio restoration and analysis," *J. New Music Research*, vol. 30, no. 4, pp. 323–338, 2001.

[2] P. Esquef, M. Karjalainen, and V. Valimaki, "Detection of clicks in audio signals using warped linear prediction," in *Proc. 14th IEEE Int. Conf. on Digital Signal Processing*, Santorini, Greece, 2002, pp. 1085–1088.

[3] H. Lin and S. Godsill, "The multi-channel ar model for real-time audio restoration," in *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, US, 2005, pp. 335–338.

[4] G. Cabras, S. Canazza, P. Montessoro, and R. Rinaldo, "he restoration of single channel audio recordings based on non-negative matrix factorization and perceptual suppression rule," in *Proc. 13th Int. Conf. Digital Audio Effects DAFx*, Graz, Austria, 2010, pp. 458–465.

[5] S. Godsill and P. Rayner, *Digital Audio Restoration A Statistical Model Based Approach*. Springer-Verlag, 1998.

[6] D. Lee and H. Seung, "Algorithms for non-negative matrix factorization," *in Advances in NIPS.*, pp. 556–562, 2000.

[7] P. Smaragdis and J. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, US, 2003.

[8] G. Cabras, S. Canazza, P. Montessoro, and R. Rinaldo, "The restoration of low-quality audio recordings based on non-negative matrix factorization and perceptual assessment by means of the ebu mushra test method," in *Proc. of ACM Multimedia International Conference*, Firenze, Italy, 2010, pp. 19–24.

[9] N. Bertin, R. Badeau, and E. Vincent, "Enforcing harmonicity and smoothness in bayesian nonnegative matrix factorization applied to polyphonic music transcription," *IEEE Trans. Audio, Speech, Lang. Processing*, vol. 18, no. 3, pp. 538–549, 2010.

[10] J. Carabias, T. Virtanen, P. Vera, N. Ruiz, and F. Canadas, "Musical instrument sound multi-excitation model for non-negative spectrogram factorization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 6, pp. 1144–1158, 2011.

[11] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "Rwc music database: Music genre database and musical instrument sound database," in *Proceedings of the 4th International Conference on Music Information Retrieval*, 2003, pp. 229–230.

[12] http://bedroomproducersblog.com/2012/04/02/free-vinyl-noises-sample-pack-released-by-mad-ep.

[13] http://www.musicradar.com/news/tech/sampleradar-243-free-vinyl-style-samples-277010.

[14] http://daviddas.com/2011/01/vinyl-record-samples-for-free-download/.

[15] http://grillobeats.com/blog/downloads/samples-vinyl/.

[16] http://www.thecontrolcentre.com/diamondsanddust.htm.

[17] http://www.partnersinrhyme.com/blog/public-domain-vinyl-record-hiss-pop-crackle/.

[18] T. Virtanen and A. Klapuri, "Analysis of polyphonic audio using source-filter model and non-negative matrix factorizationn," *Advances in Models for Acoustic Processing, Neural Information Processing Systems Workshop*, 2006.

[19] Paco de Lucia plays Manuel de Falla, Record company: Polygram Iberica S.A, 1978.

[20] Concerto De Aranjuez, Record company: Polygram Iberica S.A, 1978.

[21] Paco de Lucia Antologia, Record company: Polygram Iberica S.A, 1995.

[22] Wavelab 6, Audio Editing and Mastering Suite from Steinberg. http://www.steinberg.net/en/products/wavelab/why_wavelab.html.

[23] E. Vincent, C. Févotte, and R. Gribonval, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.