Comparison of different audio restoration methods based on frequency and time domains with applications on electronic music repertoire

Sergio Canazza, Giovanni De Poli, Gian Antonio Mian, Alessandro Scarpa

Dept. of Electronics and Informatics University of Padova - V. Gradenigo 6/a - 35100 Pd - Italy
{canazza, depoli, mian, skarpa}@dei.unipd.it

Abstract
In recent years the evolution of digital technologies for the coding of audio signals has lead to new perspectives in the field of preservation and fruition of audio archives, even though new models for digital audio archives have not been yet developed. The most important international archives make use today of digital databases, and an increasing number of digital supports (DAT, CD-A, DVD-A). At the same time, in the field of musical philology, restored editions of electroacoustic operas have never been released: these would lead to scientific problems never addressed so far, about methodologies and algorithms. We aim at applying restoration algorithms to recordings of electronic music; this repertoire has some peculiar problems: it requires a philological analysis of the piece, as many of the synthesized sounds have acoustic properties in common with noise, both impulsive and white.

I. INTRODUCTION

This work presents the results of an experiment aimed to evaluate the quality of different audio restoration algorithms based on different methods, based both time and frequency domain.

A look at the most interesting ideas that have developed from the eighties onwards has revealed that the methodologies and techniques for the compensation and removal of signal deteriorations adopted up to now have been conceived, above all, on the basis of sound sources of an acoustic nature: voices, musical instruments, live or recorded artistic performances, etc. Very rare are, instead, the contributions in the field of the preservation and restoration of tapes of electronic music. A substantial patrimony of compositions for magnetic tapes still lies almost forgotten and awaits to be saved and transferred onto more permanent and long-lasting carriers.

These tapes, which were mostly recorded over the first twenty years after World War II and which are held in the archives of many musical institutes and radios, need to be safeguarded with specific measures of preservation, without which the works they contain – which exist as recordings only – would be irreparably lost.

Besides lying in the impalpable and purely artistic values which are materialized thanks to the recording instruments of the time, their documentary quality stands out from that of other musical reproductions for the specific knowledge on the technology of electronic music that is an integral part of the composition process fixed on the magnetic tapes. These are fundamental aspects to consider when making decisions on the document, on its preservation and on the process of transferring the information it contains.

The peculiarity of the audio material present in the electronic compositions on tape, the technology with which they were generated, and the cultural modifications produced by a new medium raise extra problems which cannot be reduced to the normal praxis of audio restoration.

A set of electronic-music recordings was selected from the archive of the Center of Computational Sonology (CSC) of Padova University and from the works of Bruno Maderna, preserved in the "B. Maderna" archive of Bologna University and in the archive of "Studio di Fonologia musicale" (RAI, Milano). To single out the best computational methodologies for audio restoration an experiment was made, implementing a software (in DirectX plug-in form) which uses the methods presented in Section II.

In commercial audio-restoration software, adaptations of the spectrally-based techniques applied in speech processing have been applied. They are based upon spectral weighting, in which individual spectral components are weighted according to expected noise and signal components. Such techniques can be viewed as finite block-size approximations to frequency domain Wiener filtering. As a result of these approximations (necessary to follow the time-varying nature of the useful signal) undesirable distortions can occur, the most notable being known as "musical noise" in which statistical fluctuations in the frequency components of noise lead to random tonal artifacts in the processed signal. Various techniques have been applied to mask or eliminate these distortions.

In our implementation different version of Wiener filtering and the Ephraim and Malah suppression rule were introduced. Moreover, we developed some algorithms based
on psychoacoustic models. This task requires transforming the audio signal from an “outer” to an “inner” representation that is to resort to a representation that takes into account how the human ear perceives the sound. The combination of the psychoacoustic model and frequency-domain algorithms permits to define a promising restoration methodology.

Time domain methods are also emerging in commercial software because of better quality produced. The main advantage is the absence of musical noise. On the other hand, these methods are computationally highly intensive compared with frequency-domain ones.

The software developed (presented in Section III), implementing these algorithms, permits to compare the different methodologies in an objective way: in fact, it is possible to use the same software environment to do restoration. In this way, is feasible to compare the results of the different restoration methods. In Section II, a short overview of the most used audio restoration methods will be presented. Section IV will be devoted to a detailed description of a time-frequency analysis of the restored stimuli in order to characterize the different algorithms used.

II. RESTORATION METHODS

A. Frequency-domain methods

In commercial audio-restoration software adaptations of the spectrally-based techniques applied in speech processing are used. They are based upon spectral weighting [3], in which individual spectral components are weighted according to expected noise and signal components. Such techniques can be viewed as finite block-size approximations to frequency domain Wiener filtering. As a result of these approximations (necessary to follow the time-varying nature of the useful signal) undesirable distortions can occur, the most notable being known as “musical noise” in which statistical fluctuations in the frequency components of noise lead to random tonal artifacts in the processed signal. Various techniques have been applied to mask or eliminate these distortions. The algorithm that gave the best results, within the explored techniques, was the Ephraim and Malah Suppression Rule (EMSR) [5]-[6]. With this method, the “musical noise” artifact is eliminated without bringing distortion into the recorded signal even if the noise is only poorly stationary and without using a crude overestimation of the noise average spectrum. The attenuation to be applied to the Short Time Fourier Transform coefficients can be expressed as the time and frequency dependent spectral gain $G(p, f)$. $G(p, f)$ depends on two parameters, $R_{\text{post}}$ and $R_{\text{prior}}$ evaluated at each frame $p$. $R_{\text{post}}$ (“a posteriori” Signal-to-Noise Ratio) is a local estimate of the Signal-to-Noise Ratio (SNR) computed from the data in the current short-time frame. $R_{\text{prior}}$ (a “priori” Signal-to-Noise Ratio) represents the information on the unknown spectrum magnitude gathered from previous frames.

Special consideration was paid to the perceptually relevant characteristics of the signal. This task requires to transform the audio signal from an “outer” to “inner” representation, that is to resort to a representation that takes into account how the human ear perceives the sound. For this purpose, according to Beerends and Stemberdinck model [2], an outer to inner ear transformation is applied to the short time power spectral density of the audio signal.

1) Noise filtering

To cope with the non-stationarity of the audio signal the Wiener filter is time-varying and is based upon short time Fourier processing (see Fig. 1).

Let $Y(p, f)$ denote the Short Time Fourier Transform (STFT) of $y(n)$, where $p = Ln$ is the analysis time index and $f = kF_s/N$ the frequency, with $F_s$ the sampling frequency, $N$ the window length and $k = 0, 1, ..., N − 1$. The method basically consists in applying a time-varying attenuation $G(p, f)$, with $0 \leq G \leq 1$, to the short time spectrum of the noisy signal:

$$
|\hat{X}(p, f)| = G(p, f) \cdot |Y(p, f)|
$$

to obtain the restored signal $\hat{x}(n)$.

According to Ephraim and Malaghi [6], the gain $G(p, f)$, $f = f_i$, is calculated for each frame as:

$$
G(f) = \frac{\sqrt{\pi}}{2} \left[ \frac{1}{1 + R_{\text{post}}(f)} \right] \left[ \frac{R_{\text{prior}}(f)}{1 + R_{\text{prior}}(f)} \right].
$$

$$
\mathcal{M} \left[ 1 + R_{\text{post}}(f) \right] \left[ \frac{R_{\text{prior}}(f)}{1 + R_{\text{prior}}(f)} \right]
$$

where $\mathcal{M}[]$ is the hypergeometric function.

The values of $R_{\text{post}}(f)$ and $R_{\text{prior}}(f)$ correspond to “a posteriori” and “a priori” estimates of the SNR at frequency $f$. The first term corresponds to a local estimate of the SNR, that is to an evaluation done on the basis of the current frame only. The second term estimates the SNR via a convex combination of the local SNR estimate, multiplied by $(1 - \alpha)$ ($0 \leq \alpha \leq 1$), and the estimate gathered from previous
frames, multiplied by $\alpha$. The factor $\alpha$ has an important meaning because if $\alpha \approx 1$ there is a greater contribution of the past frames, while if $\alpha \approx 0$ there is a greater contribution of the actual frame. When the signal is very noisy, the value of $\alpha$ should be near to 1 because there must be a great attenuation, at contrary, for high SNR values, $\alpha$ should be near to 0. In [5], Cappé suggests the value $\alpha = 0.98$.

A variant of the method, in the sequel denoted as $W_i$, takes into account this fact: if $R_{\text{post}}(p, f) > 0$, $\alpha$ is set to 0.98; else, if $R_{\text{post}}(p, f) + R_{\text{post}}(p - 1, f) < 0$, then $\alpha = 0$, while if $R_{\text{post}}(p, f) < 0$ and $R_{\text{post}}(p - 1, f) > 0$, then $G(p, f) = G(p - 1, f)$. This provision slightly improves the behavior of the algorithm in the attacks, i.e., in transitions from noise to signal.

2) Psychoacoustic model

To filter the noise in a perceptually meaningful way, it is necessary to transform the audio signal from an “outer” to “inner” representation, i.e., into a representation that takes into account how the sound waves are perceived by the auditory system. The device used is the Beerends and Steemerink model [2], sketched in Fig. 2. The signal $x(n)$ is first windowed by the $w(n)$ window and transformed in the frequency domain. The short time spectral power is transformed from Hertz ($f$) to Bark ($z$) scale, bandlimited and spread both in time and frequency.

![Fig 2: The audio signal transformation from “outer” to “inner” representation.](image)

As a result, the outer frequency domain representation $Y(p, f) = X(p, f) + D(p, f)$, with $X$ and $D$ signal and noise spectrum estimates, is transformed into the internal representation $Y(p, z) \approx \bar{X}(p, z) + \bar{D}(p, z)$, defined in the Bark domain, bandlimited and processed taking into account the spreading both in time and frequency. Finally, the $R_{\text{pre}}(p, z)$ and $R_{\text{post}}(p, z)$ terms are calculated according to the inner representation and the gain $G(p, z)$ is derived.

B. Time-domain methods

Substantially different from previous framework, it is possible to work completely in time domain. Here the Niedzwiecki and Cisowski procedure is considered. Details can be found in [8], and improvements in [1]. The signal is modeled as a $p$ order time varying autoregressive (AR) model:

$$s(t + 1) = \sum_{i=1}^{p} a_i(t) s(t - i + 1) + e(t)$$

driven by the gaussian zero-mean white noise sequence $e(t)$ with variance $\sigma_e^2$. The time evolution of the time varying coefficients $a_i(t)$ is modeled by the random walk model:

$$a_i(t + 1) = a_i(t) + w_i(t), \quad i = 1, ..., p$$

with $w_i(t)$ zero mean gaussian white processes of variance $\sigma_w^2$ mutually uncorrelated, i.e., $E[w_i(t) w_j(t)] = 0$ for $i \neq j$, and independent of $e(t)$. It is important to note that this model can be extended with impulsive noise detection and reduction. However in this paper we focus only on broadband noise.

In [8] is developed a procedure for detection/tracking/restoration of signal at the same time, as a non-linear filtering problem using Extended Kalman Filter (EKF) theory. Moreover in [1] are explained some improvements such as forward/backward filtering and variable memory tracking (VFF algorithm).

III. RESTORATION TOOL

To single out the best computational methodologies for audio restoration, an experiment was made, implementing a software (in DirectX plug-in form) which uses the methods presented in Section II (see fig. 3).

![Fig. 3a: The user interface of CSC Restoration Tool. The restoration, the setting of $\alpha$ value and the noise print modification are performed in real time.](image)
Usually, different algorithms are implemented in different software, with different user interfaces and the parameters are used in different way (and not documented). Moreover, often, in commercial products, the Software House doesn’t indicate in detail the algorithm implemented. So, the comparison carried out with commercial products, test the software quality (i.e. the implementation quality), not the algorithm effectiveness. On the contrary, our software permits to compare different algorithms in objective way: in fact, it is possible to use the same software environment to operate the restoration. In this way, is feasible to compare the different methods.

The filters implemented are: Wiener (standard), Ephraim and Malah Suppression Rule (EMSR), a number of EMSR variations (like $W_3$ described in Section II), and algorithms based on psychoacoustic model, that use different noise suppression rule (Wiener, EMSR, $W_3$). In restoration based on EMSR (and its variations), the user can set the $\alpha$ value accordingly to the particular signal considered. Moreover, the tool permits to modify (in real time) the noise print estimated, by an ‘equalizer’ in bark scale.

For the Extended Kalman Filter, the entire procedure of [8] is implemented, with the possibility to perform dehiss and declick with a single pass. The threshold for click detection and the noise variance can be modified in real time. Furthermore other improvements were added (as in [1]) especially forward/backward filtering and adaptive tracking of music signal ($VFF$ algorithm and decimation of tracking).

IV. VALIDATION

To validate the model, several recordings (with different SNR) were restored using the CSC Restoration Tool described above. As an example, let us consider the restoration of an excerpts taken from “Komposition für Oboe, Kammerensemble und Tonband” (1962) of B. Maderna (from archive of Bologna University). Fig. 4 shows the spectrograms of the original recordings (sampled at 44.1 kHz and 16 bit). Fig. 5, 6, 7, 8 show the spectrograms of the same excerpt restored with, respectively, Wiener filter (fig. 5), $W_3$ algorithm (fig. 6), psychoacoustic model based on Wiener filter (fig. 7) and EMSR algorithm (fig. 8). The parameters used to control the different algorithms were subjectively tuned to obtain the best tradeoff between noise-removal and music-signal-preservation.
From the comparison among these frequency-based methods, the following situation is outlined. The $W_2$ algorithm results the best one: with an appropriate setup of $\alpha$ parameter, window size and window overlap, it presents a great noise reduction without to cause audible distortion.

The EMSR presents a slight ‘musical noise’ (annoying only with very noisy signal), but the $\alpha$ parameter must be set to obtain the best tradeoff between noise reduction and transients preservation.

The standard implementation of Wiener filter has only educational value, given his analytical simplicity.

The perceptual filters have a common characteristic: where SNR $\approx 0$ the noise is correctly reduced; on the contrary, where SNR $> 0$, there is too much residual noise (it is considered erroneously masked by the filters). In particular, the perceptual filter based on EMSR shows a low-pass effect. Probably, this effect is due to the combined influences of spreading (both in time and frequency) and of “a priori” estimates of the SNR ($R_{pri}$).

In the presence of a great noise amount, the ‘musical noise’ is present also using the $W_2$ filter. In this case, an appropriate setup of noise print overestimation (in particular, increasing the overestimation values in high frequency region), can reduce this artifact.

The difference among time method algorithms are very subtle and not graphically perceivable. Therefore, it is necessary to listen to the examples, on-line at

By means perceptual tests, we deduced that the best quality reached is obtained using Extended Kalman Filter with VFF tracking and scalar combination of forward/backward passes for high SNR values (say 35dB). For lower SNR values $W_2$ is more appropriate.

V. CONCLUSIONS

Audio materials are recorded in various supports in which a rapid degradation of the information occurs. For the preservation, restoration and handling of a huge and often badly preserved audio heritage, we need to develop methodologies which allow to classify degradation of audio material, define a restoration protocol on the basis of the kind of degradation, and to project methodologies for preservation and handling audio archives.

We aim at applying restoration algorithms to recordings of electronic music; this repertoire has some peculiar problems: it requires a philiological analysis of the piece, as many of the synthesized sounds have acoustic properties in common with noise, both impulsive and white.

In this paper, we have presented the results of an experiment aimed to evaluate the quality of different audio restoration algorithms based on different Short Time Spectral Attenuation methods and on Extended Kalman Filter. For this purpose, an experiment was made, implementing a software in DirectX plug-in form. The software, implementing several algorithms, has permitted to compare the different methodologies in an objective way.

A global analysis can be carried out. The perceptual filters don’t seem suitable to noisy signal with low SNR, because they leave an annoying residual noise. The $W_2$ filter results the best one in the frequency domain: with an appropriate setup of parameters, it presents a great noise reduction without to cause audible distortion. Generally, the best choice is EKF because of the absence of musical noise and the inaudible low-pass effect; on the contrary this filter requires a long training to fine-tune its parameters.

VI. REFERENCES


