REAL TIME COMPARISON OF AUDIO RESTORATION METHODS BASED ON SHORT TIME SPECTRAL ATTENUATION

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
http://www.dei.unipd.it/~musica

ABSTRACT

This paper presents the results of an experiment aimed to evaluate the quality of different audio restoration algorithms based on different Short Time Spectral Attenuation methods. To single out the best computational methodologies for audio restoration an experiment was made, implementing a software (in DirectX plug-in form) which uses different algorithms.

The software, working in real time, permits to compare the different algorithms in a objective way: in fact, it is possible to use the same software environment to operate the restorations with different algorithms. In this way, is feasible to compare different methods.

In the paper we will first shortly overview the most used audio restoration methods and, in particular, the algorithms implemented in our software. Then we will present a time-frequency analysis of the restored stimuli to show the main advantages and drawbacks of the different algorithms used.

1. INTRODUCTION

Audio effects, normally, tends to modify sound to give it a particular character or atmosphere; audio restoration, instead, tends to recover the “original” sound after it was deteriorated by aging of support or bad conservation [1]. In recent years the evolution of digital technologies for the processing 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 electro-acoustic operas have never been released: these would lead to scientific problems never addressed so far, about methodologies and algorithms.

While most studies on restoration deal with classical and pop music restoration, little attention has been paid to electro-acoustic 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 [2, 3]. Few algorithms comparison has been done so far.

This work presents a system to evaluate the quality of different audio restoration algorithms based on different Short Time Spectral Attenuation methods. Some comparisons on a set of electro-music recordings will be presented. 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 2.

In commercial audio-restoration software, adaptations of the spectrally-based techniques used in speech processing have been applied [4]. 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.

A real time comparison of different algorithms and setting is not possible in commercial software; instead, it would very useful in order to evaluate which algorithms and settings are more suitable to the music being restoring both for educational and production use. The software developed (presented in section 3), 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 and settings. In section 2, a short overview of the most used audio restoration methods will be presented. Section 4 will be devoted to a detailed description of a time-frequency analysis of the restored stimuli in order to characterize the different algorithms used.

This work was supported by CNR - "Progetto Finalizzato Beni Culturali".

2. 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 [5], 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) [6, 7]. 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_{post} and R_{prio} evaluated at each frame *p*. R_{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_{prio} (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 Stemerdink model [8], an outer to inner ear transformation is applied to the short time power spectral density of the audio signal.

2.2. 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 $Y(p,f_k)$ denote the Short Time Fourier Transform (STFT) of $y(n)$, where $p=Ln$ is the analysis time index and $f_k=kF_s/N$ 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 $G(p,f_k)$, with $0 \le G \le 1$, to the short time spectrum of the noisy signal:

$$\left|\hat{X}(p, f_k)\right| = G(p, f_k) \cdot \left|Y(p, f_k)\right|$$

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

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

$$G(f) = \frac{\sqrt{\pi}}{2}\sqrt{\left(\frac{1}{1+R_{post}(f)}\right)\left(\frac{R_{prio}(f)}{1+R_{prio}(f)}\right)} \cdot$$
$$\cdot M\left[\left(1+R_{post}(f)\right)\left(\frac{R_{prio}(f)}{1+R_{prio}(f)}\right)\right]$$

where *M[·]* is the hypergeometric function.

The values of $R_{post}(f)$ and $R_{prio}(f)$ 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-α) ($0 \le \alpha \le 1$), and the estimate gathered from previous frames, multiplied by α. The factor α has an important meaning because if $\alpha \approx 1$ there is a greater contribution of the past frames, while if $\alpha \cong 0$ there is a greater contribution of the actual frame. When the signal is very noisy, the value of α should be near to 1 because there must be a great attenuation, at contrary, for high SNR values, α should be near to 0. In [7], Cappé suggests the value α = 0.98.

A variant of the method, in the sequel denoted as W_2 takes into account this fact: if $R_{post}(p,f) > 0$, α is set to 0.98; else, if $R_{post}(p,f)$ and $R_{post}(p-1,f) < 0$, then α= 0, while if $R_{post}(p,f) < 0$ and $R_{post}(p-1,f) > 0$, 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.

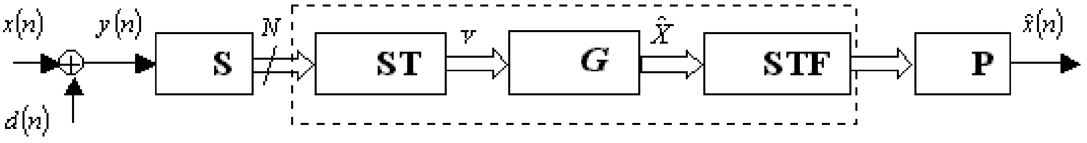


Fig 1: The model used for restoration: *x(n)* is the "noise free" signal, *d(n)* the noise, *y(n)* the degraded signal, and $\hat{x}(n)$ the restored signal; S/P and P/S represent series-to-parallel and parallel-to-series converters; **G** = *diag*{*G(p,f)*}.

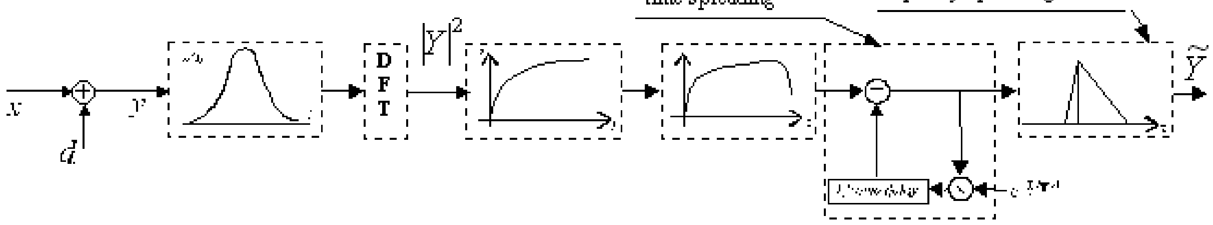
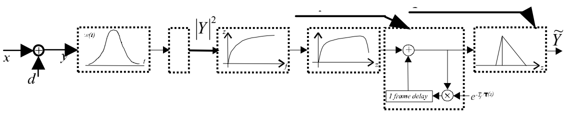


Fig 2: The audio signal transformation from "outer" to "inner" representation.

2.3. Psychoacoustic model

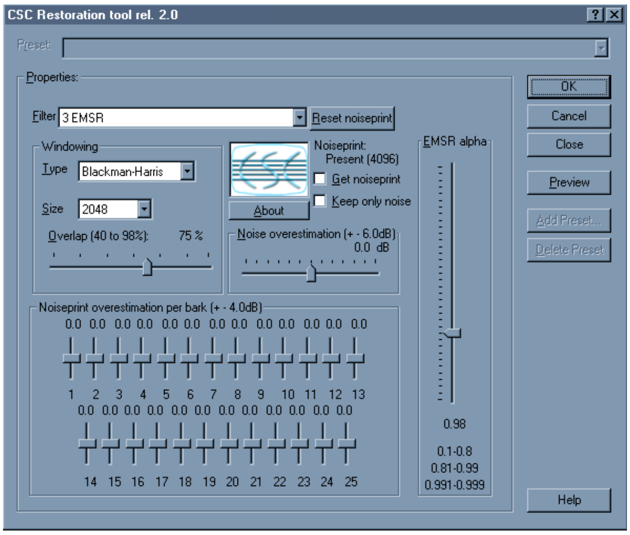
To filter the noise in a perceptually meaningful way, it is necessary to transform the audio signal from an "outer" to "inner" representation, i.e., into a representation that takes into account how the sound waves are perceived by the auditory system. The device used is the Beerends and Stemerdink model [8], 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.



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 $\widetilde{Y}(p,z) \approx \widetilde{X}(p,z) + \widetilde{D}(p,z)$, defined in the Bark domain, bandlimited and processed taking into account the spreading both in time and frequency. Finally, the $\widetilde{R}_{prio}(p,z)$ and $\widetilde{R}_{post}(p,z)$ terms are calculated according to the inner representation and the gain $\widetilde{G}(p,z)$ is derived.

3. RESTORATION TOOL

To single out the best computational methodologies for audio restoration, a software tool was developed (in DirectX plug-in form) which uses the methods presented in section 2 (see fig. 3). A demo version is downloadable at http://www.dei.unipd.it/~musica. The Direct X technology allows to integrate such modules in all audio editor (and similar programs) running on Windows platform.



CSC Restoration tool rel. 2.0

Preset:

Properties:

Filter: 3 EMSR

Windowing

Type: Blackman-Harris

Size: 2048

Overlap (40 to 98%): 75 %

Noiseprint:
Present (4096)

□ Get noiseprint

□ Keep only noise

Noise overestimation (+ - 6.0dB) 0.0 dB

Noiseprint overestimation per bark (+ - 4.0dB)

1	2	3	4	5	6	7	8	9	10	11	12	13
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

14	15	16	17	18	19	20	21	22	23	24	25
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

EMSR alpha: 0.98

0.1-0.8
0.81-0.99
0.991-0.999

Fig. 3: The user interface of CSC Restoration Tool. The restoration, the setting of α value and the noise print modification are performed in real time.

Usually, different algorithms are implemented in different software, with different user interfaces and the parameters are used in different way (and don't documented). Moreover, often, in commercial products, the Software House don'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_2 described in section 2.2), and algorithms based on psychoacoustic model, that use different noise suppression rule (Wiener, EMSR, W_2). In restoration based on EMSR (and his variations), the user can set the α 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.

4. 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 shows the spectrograms of the same excerpt restored with, respectively, Wiener filter (fig. 5), W_2 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 setup to obtain the best tradeoff between noise-removal and music-signal-preservation.

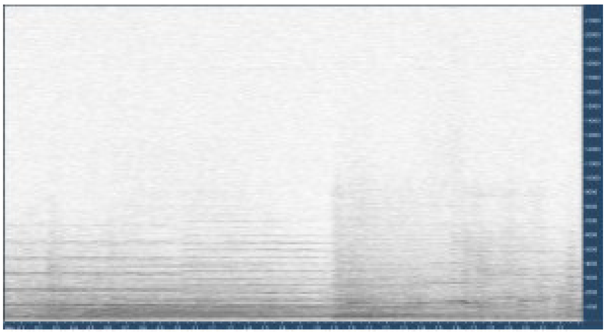


Fig. 4: Spectrogram of "Komposition für Oboe, Kammerensemble und Tonband " (1962) of B. Maderna, from 7'35'' to 7'38''.

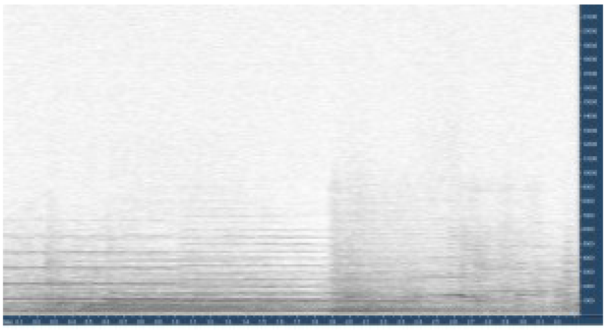


Fig. 5: Spectrogram of "Komposition für Oboe, Kammerensemble und Tonband " (1962) of B. Maderna, from 7'35'' to 7'38''. Restored with Wiener filter.

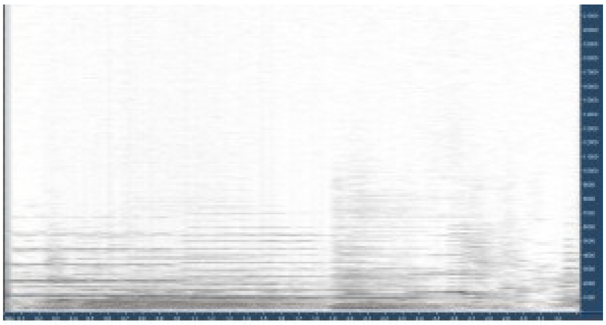


Fig. 6: Spectrogram of "Komposition für Oboe, Kammerensemble und Tonband " (1962) of B. Maderna, from 7'35'' to 7'38''. Restored with EMSR algorithm.

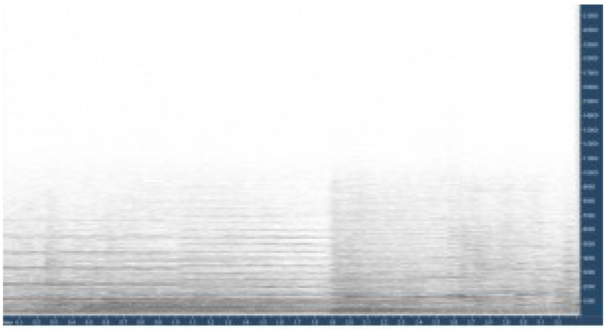


Fig. 7: Spectrogram of "Komposition für Oboe, Kammerensemble und Tonband " (1962) of B. Maderna, from 7'35'' to 7'38''. Restored with psychoacoustic model (with Beerends and Stemerdink model).

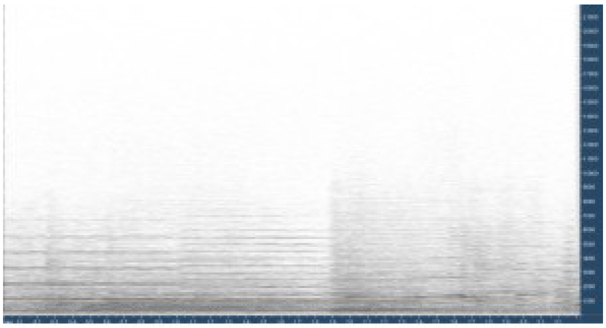


Fig. 8: Spectrogram of "Komposition für Oboe, Kammerensemble und Tonband " (1962) of B. Maderna, from 7'35'' to 7'38''. Restored with psychoacoustic model (with Beerends and Stemerdink model, based on EMSR).

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 α 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 α 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 ≈ 0 the noise is correctly reduced; on the contrary, where SNR >> 0, there is too much residual noise (erroneously, it is considered 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_{prio}).

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.

5. CONCLUSION

Audio materials are recorded on 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 philological 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 presented the results of an experiment aimed to evaluate the quality of different audio restoration algorithms based on different Short Time Spectral Attenuation methods. For this purpose, a real time software tool, realized in DirectX plug-in form, was developed. The software, implementing several algorithms, permits to compare the different methodologies in real time.

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: with an appropriate setup of parameters, it presents a great noise reduction without to cause audible distortion.

6. REFERENCES

[1] D. Schueller, "The ethics of preservation, restoration and re-issues of historical sound", *J. Audio Eng. Soc.*, 39(12), pp. 1014-1016, 1991.

[2] S. Canazza, G. Coraddu, G. De Poli, G. A. Mian, "Objective and subjective comparison of audio restoration methods", *Journal of New Music Research*, in press.

[3] A. Bari, S. Canazza, G. De Poli, G. A. Mian, "Toward a methodology for the restoration of electro-acoustic music", *Journal of New Music Research*, in press.

[4] S. Godsill, P. Rayner, O. Cappé, "Digital audio restoration". In *Applications of digital signal processing to audio and acoustics*, Kahrs - Karlheinz Brandeburg (ed.), Kluwer Academic Publishers, 1998.

[5] S.F. Boll, A.V. Oppenheim, "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-27(2), April 1979.

[6] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator", *IEEE Trans. Acoustics, Speech and Signal Processing*, 21(6) pp. 1109-1121, 1984.

[7] O. Cappé, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor". *IEEE Trans. Speech and Audio Processing*, vol. 2(2), pp. 345-349, 1994.

[8] J. G. Beerends, J. A. Stemerdink, "A Perceptual Audio Quality Measure Based on Psychoacustic Sound Representation", *J. Audio Eng. Soc.*, 40(12), pp. 963-978, 1992.