Int J Digit Libr (2012) 12:121–135
DOI 10.1007/s00799-012-0088-x

The digital curation of ethnic music audio archives: from preservation to restoration

Preserving a multicultural society

Sergio Canazza

Published online: 22 March 2012

Abstract In the sound archive field, a long-term maintenance of the collective human memory in its original form is not sustainable. All physical carriers are subject to degradation and the information stored on such carriers is bound to vanish. Only a *re-mediation* of the original documents can prevent precious knowledge from being permanently lost. In particular, ethnic music audio documents are often recorded on non-professional carriers by means of amateur recording system, or, more in general, in fieldwork circumstances. Thus, the preservation of the carrier and the restoration of the audio signal are crucial to avoid the permanent loss of the musical heritage, which is already heavily corrupted. This article describes the protocols defined, the processes undertaken, the results ascertained from several audio documents preservation/restoration projects carried out in the ethnic music field, and the techniques used. In particular: (i) a number of recommendations are given for the re-recording process, and (ii) an audio restoration environment (constituted by three audio restoration tools), developed using the VST plug-in architecture and optimized for different audio carriers (cylinders, shellac discs, tapes) is detailed. The experimental results and the perceptual assessment presented show the effectiveness of the restoration environment developed by the author.

Keywords Audio archives · Ethnic music · Audio restoration · Historical audio documents

S. Canazza (✉)
Sound and Music Computing Group, Department of Information Engineering, University of Padua, Padua, Italy
e-mail: sergio.canazza@unipd.it

1 Introduction

The analysis of Western music has been developed almost exclusively on the basis of written scores, which represent musical performance models, rather than the performance itself.

This work refers to the audio documents preservation/restoration related to ethnic music, a genre characterized by cultural processes rather than abstract musical types related not only to the feudal, capitalist and some oriental societies (e.g., Union of Soviet Socialist Republics, People's Republic of China) but also to "primitive" and "popular cultures" [1]. In ethno-musicology, music is part of a cultural context and social life: the audio document contains information about the whole context in which the document was produced. These audio documents are affected by the problems of multiple carrier[1] and formats.

Ethno-music archives and cultural institutions begin to show a keen awareness both for the opportunities offered by the new media to improve and spread their documentary heritage, and for uncontrolled duplication and manipulation risks arising from transfer onto digital carriers and to on-line access. There have been several EU research projects on *digital curation*[2] of audio documents in different contexts and from different points of view (IST, Culture2000, Mediaplus, Interreg, eContentplus, etc.), but very few of them are focused on ethnic music. Moreover, as it is well known, the multidisciplinary research in a multicultural society preservation—of which the (carrier and signal) restoration of audio documents

[1] E.g., in analogue domain: wax cylinders, sonofils, discs, tapes, and cassettes; in digital domain: magnetic tapes and optical discs.

[2] Digital curation is the process of establishing and developing long term repositories of digital assets for current and future reference by researchers, scientists, and historians, and scholars generally [2].

is an outstanding example—fits perfectly in the Sound and Music Computing roadmap [3] (challenge 3) as defined in the scientific community.

Despite the youthfulness of the musical recording technologies and the content, their preservation is already becoming an important issue to avoid the loss of precious cultural heritage recordings. These documents have a shorter life span compared to other cultural heritage materials and their maintenance and restoration introduce novel problems requiring different and original approaches. Ethnic music, in particular, is not based on score notation, thus the traditional Conservatories of Music are unable to "preserve" the performance praxis.

The ethnic music audio documents are particularly complex in terms of preservation, signal restoration, retrieval, and access. The ethnic-musical heritage is in danger of disappearing, of being forgotten in some public archive or, in most cases, a private collection, because of the poor quality of the material on which the audio documents were recorded and the rapid evolution of the recording formats—that make obsolete and scarcely readable many old recordings: the ethnic music recordings were often made with non-professional systems (low-quality, poorly aligned and maintained—often by technically unskilled researchers—without generally accepted standards and recording practices) or, more in general, in fieldwork circumstances (that implies a low signal to noise ratio (SNR) in the audio signal). Obviously, this has not been the destiny of music repertoires of wider interest, such as vocal or classical music, rock/pop, or jazz. In these cases, the recording companies have re-recorded most of the audio documents, particularly those of high commercial value. Unfortunately, the same has not happened to ethnic and traditional repertoires.

The ethnic music audio documents are often the only witnesses of disappeared oral cultures. Their preservation requires procedures, detailed for every type of carrier, able to preserve not only the recorded audio but also all the metadata and contextual information which can be useful for the musicological research. In this sense, this field of study is particularly important to preserve the document and to restore the audio signal in a musicologically correct way. It must be noted that in the past the lack of musicological severity and attention to source consistency has already produced several gross mistakes in the A/D transfer of musical audio archives, producing yet more sources which often feature disfiguring transmission errors. Interesting examples can be retrieved in [4]. These actions presuppose the presence of several highly skilled experts, since the damages of a bad preservation or an inadequate (carrier and) signal restoration are usually irreversible. Preservation and restoration activities should be planned and performed within agreed international standards and procedures and carried out within institutional environments. This is a challenge in which signal processing, sound and music computing, and ICT research can contribute significantly.

By studying examples from this field, a broad range of challenging problems has been discovered and analyzed.

For what concerns audio memories, preservation is divided into passive preservation, meant to defend the carrier from external agents without altering the structure, and active preservation, which involves data transfer onto new media. Active preservation allows us to copy, handle and store the audio documents in virtual spaces that can be remotely accessed by large communities. Digitization is necessary to prevent the documents from disappearing, and it is desirable because it allows to distribute them on a wide scale. The commingling of a technical and scientific formation with historic-philological and philosophical knowledge also becomes essential for preservative re-recording operations, which do not completely coincide with pure A/D transfer, as is, unfortunately, often thought. Three examples will be made, in different music genres.

(1) *Stria* (John Chowning, 1977): in the CCRMA version (four-channels), there was a signal discontinuity in the D/A conversion of the data at 6′ 29″ (389″) from the original computation that was unintended. This caused a sudden change of timbre and a consequent click: "[T]he PDP-10 burped!" [5]. This imperfection in the computation emerges slightly, but very clearly indeed, in the audio source. John Chowning did not re-compute the section to eliminate this problem. He rather learned to accept it "as one does a birth mark or beauty mark on ones skin…noticeable but of no substantive consequence" [5]. The faulty, imperfect, and therefore fascinating four-channel version is the version that John Chowning now uses to play during the concerts. Conversely, in the commercial version (CD Wergo, WER 2012-50) this "burp" is missing. The audio is truncated exactly at that point (6′ 29″) with a fade out to the following section. This is an example of a lack of the philological attention during the re-recording process. For a detailed description of this transmission error see [5].
(2) *Y entonces comprendió* (Luigi Nono 1970): it is a four-channel music work. Luigi Nono produced also a stereophonic version in a four-channel tape (A, B, C, D), mixing the original four-channel tape (1, 2, 3, 4): $A = 1+3$; $B = 1+3$; $C = 2+4$; $D = 2+4$. The Stereo Long Playing Deutsche Grammophon DGG 2530436 (stereo: X, Y) was mixed: $X = A + C$; $Y = B + D$. In this way, a stereo version is reduced to a monophonic version, because of a transmission error.
(3) The commercial audio discs, dating from 1894, have been recorded following a large set of different carriers and encodings. As for their physical composition, audio

> discs can range from fragile forms such as rubber (the earliest disc recordings), acetate, or lacquer (sometimes with glass, aluminum, or cardboard backings), to more-durable shellac and vinyl discs and the metal masters used to stamp commercial discs. The distinct characteristics of each disc type require different techniques, often highly specialized, to coax the sound from the carrier. So, in this case several choices (in relation with the phonographic disc history) are necessary to optimize the extraction of the audio signal from the original carrier: the pick-up arm, the cartridge, the stylus, the speed, and the replay equalization are all factors that influence the result of the re-recording process.

This work aims at presenting the state of the art in the field of preservation and signal restoration of ethnic music audio documents considering several problems, that can be summarized in the following question: "Where does the *authenticity* of a musical audio document lie, when there are neither scores nor a univocal and well-established performance praxis?"

Considering the term authenticity, a wide variety of possible meaning appears just by looking at the different studying fields; from archaeology, philosophy, psychology, history this concept is altered, measured and directed into the specific focus fields. In this article the term authenticity will be dealt in the field of musicology, and even more specifically: the field of the musical recordings. Several sublevels of authenticity will be enumerated: in particular one of these—let us call it *playback authenticity*—is surely related with audio documents preservation/restoration. Recording music does not guarantee the identical sound when played back. All the more, it is even impossible to recreate the audio signal 100 % identical as the sounding source of the signal. Regarding the way sound is captured, there are already—human inaudible—signals omitted because of the chosen recording capacities. Recording (especially digital) can be seen as a bottle neck: the selected values of sample rate and bit rate are the maximum quality borders of the signal recorded. Irrecoverably cropped at the chosen bit rate and sample rate, the signal cannot regain its original spectrum. Nowadays standards of digitalization prescribe 24 bit and 96 kHz (or even 192 kHz) because that covers sufficiently the range of the human ear capacities. But speaking in purely physical terms is not possible to create the ultimate identical playback of a recorded sound. However trying to aim for the best playback possible contains already quite some challenges: speed determination, equalization, type of (analogue) playback machine, and so on. For correct playback of a music, the author-intended authenticity can come close to the meant sound, but simply applied or wrong interpreted author intentions (composer-made notations) are dangerous for misleading playback. For music from oral cultures, there is no author guiding opinion. So the musicologist has to make an objective decision in variable settings of playback. Aiming at the objective best playback—sounding as during recording—most analogue audio carriers need study before re-recording can proceed successfully. The multiple physical playback problems have caused that this category is not assimilated into the author intended authenticity. Enumerating some topics of the list of possible dilemmas of playback, the exact speed determination of most older analogue recordings is very difficult. The correct equalization of the sound is also very important for the appropriate proportion of the low–middle–high frequencies of the sound. With magnetic tapes great importance has to be given to the type of head which was used during the recording. Using another type of head will significantly lower the dynamic range of the music. Other phenomena involved in a correct playback are the azimuth (horizontal equilibration of the head) and the sticky shed syndrome (polyester tapes are slightly sticking because of chemical breakdown of the binder, so *cooking* is necessary).

Section 2 summarizes the most significant positions of the debate evolved since the Seventies inside the archival community on the historical faithfulness of the active preservation. Section 3 proposes a preservation protocol defined ad hoc for ethnic audio archives. The ethnic music audio documents are usually recorded with a SNR. Thus, for their appropriate fruition and/or for a suitable use of music information retrieval (MIR) techniques, it is necessary to process the audio signals by means of audio restoration techniques. Section 4 details an audio restoration environment (constituted by three audio restoration tools), developed using the VST plug-in architecture and optimized for different audio carriers (cylinders, shellac discs, tapes). Experimental results of these tools are presented in Sect. 5. In order to validate the system, two analyses were carried out, to demonstrate the effectiveness of the restoration environment:

(1) an objective comparison, measuring the gain introduced by the tools developed by the author, in comparison with some *standard* filters at the varying of the noisy signal SNR, considering 35 ethnic music documents;
(2) a perceptual assessment, using the EBU MUSHRA test protocol [6] (Sect. 6).

2 Historic overview

To date, the A/D transfer represents the only solution able to efficiently contrast the eclipse of audio documents and to guarantee content transmission. During the re-recording onto the new media, the communication forms change, thus outlining a plurality of approaches to the audio documents, which variously raise the typical philological problems related to the authenticity and interpretation of the documents.

A reconnaissance on the most significant positions of the debate evolved since the Seventies inside the archival community on historical faithfulness of the active preservation points out at least three different perspectives [7].

Storm [8] individuates two types of re-recording which are suitable from the archival point of view: (1) the sound preservation of audio history, and (2) the sound preservation of an artist. The first type of re-recording (Type I) represents a level of reproduction defined as the perpetuation of the sound of an original recording as it was initially reproduced and heard by the people of the era. The second type of re-recording (Type II) was presented by Storm as a more ambitious research objective: it is characterized by the use of a different playback equipment than the original one, with the intent of obtaining the live sound of original performers, transcending the limits of a historically faithful reproduction of the recording.

Schüller [9] and (cited in) [10] points directly towards defining a procedure which guarantees the the best quality of the re-recording of the signals, minimizing the audio processing as much as possible. He goes on to an accurate investigation of signal alterations, which he classifies in two categories: (1) intentional and (2) unintentional. The former includes recording equalization, and noise reduction systems, while the latter is further divided into two groups: (i) caused by the imperfection of the recording technique of the time (e.g., distortions), and (ii) caused by misalignment of the recording equipment (wrong speed, deviation from the vertical cutting angle in cylinders or misalignment of the recording head in magnetic tapes). The choice whether or not to compensate for these alterations reveals different re-recording strategies: (A) the recording as it was heard in its time (Storm's Audio History Type I); (B) the recording as it was produced, precisely equalized for intentional recording equalizations (1), compensated for eventual errors caused by misaligned recording equipment (2ii) and replayed on modern equipment to minimize replay distortions; (C) the recording as produced, with additional compensation for recording imperfections (2i).

Brock-Nannestad [11] examines the re-recording of acoustic phonographic recordings (pre-1925). In order to have scientific value, the re-recording requires a complete integration between the historical-critical knowledge which is external to the signal (ancillary information) and the objective knowledge which can be inferred by examining the carrier and the degradations highlighted by the analysis of the signal.

3 Preservation protocol

Starting from the positions listed in Sect. 2, let us define the preservation copy as a digital data set that groups the information carried by the audio document, considered as an artifact. It aims at preserving the documentary unity, and its bibliographic equivalent is the facsimile or the diplomatic copy. Restoration processes are allowed only when they are finalized to the carrier restoration. Differing from the Schüller position, my belief is that—in a preservation copy—only the intentional alterations should be compensated (correct equalization of the re-recording system and decoding of any possible intentional signal processing interventions). On the contrary, all the unintentional alterations (also those caused by misalignments of the recording equipment) could be compensated only at the access copy level (i.e., in a duplication of the preservation *archive* copy) these imperfections/distortions must be preserved because they witness the history of the audio document transmission.

The A/D transfer process should represent the original document characteristics as it arrived to us. According to the indications of the international archive community [12, 13]: (1) the recording is transferred from the original carrier; (2) if necessary, the carrier is cleaned and restored so as to repair any climatic degradations which may compromise the quality of the signal; (3) re-recording equipment is chosen among the current professional equipment available in order not to introduce further distortions, respecting the original mechanical analogies; (4) the sampling frequency and bit rate must be chosen in respect of the archival sound record standard (at least, 48 kHz/24 bit, following the slogan: *the worse the signal, the higher the resolution*[3]); (5) the digital audio file format should support high resolution, and should be transparent with simple coding schemes, without data reduction.

The process of active preservation produces a particularly large and various set of digital documents, which are made up of the audio signal, the metadata and the contextual information (the term metadata indicates content-dependent information that can be automatically extracted by the audio signal; contextual information indicates the additional content-independent information).

In order to preserve the documentary unity it is therefore necessary to digitize contextual information, which is included in the original document and the metadata which comes out from the transfer process: the information written on edition containers, labels and other attachments should be stored with the preservation copy as static images, as well as the photos of clearly visible carrier corruptions. A video of the carrier playing—synchronized with the audio signal—ensures the preservation of the information on the carrier (physical conditions, presence of intentional alterations, corruptions, graphical signs). The video file should be stored with the preservation copy. The selected resolution and the compression factor must at least allow to locate the signs and corruptions of the support. In the author's experience, the resolution of 320 × 240 pixels and a DV JPEG compression,

[3] George Brock-Nannestad: a personal oral communication.

with no more of 65 % of quality is considered a satisfactory tradeoff. In this way, it is possible to use the metadata extraction methods to describe the carrier corruptions presented in [14].

Finally, a descriptive sheet (consistent with the database of the audio archive) with all the information on the A/D equipment used and the original document, should be stored in the preservation copy in order to simplify the access to data and to avoid concurrency problems.

4 Access copy

The ethnic music audio documents are usually recorded with a SNR. Thus, for their appropriate fruition and/or for a suitable use of MIR techniques it is necessary to process the audio signals by means of audio restoration techniques. This section presents an audio restoration environment (constituted by three audio restoration tools), developed using the VST plug-in architecture and optimized for different audio carriers (cylinders, shellac discs, tapes).

Differing from the Schüller position, I propose four different approaches that can be adopted in a combined way with audio restoration algorithms, in accordance with the final purposes of the access copy:

- Documental approach: in this case, the de-noise algorithms only concern the cases in which the internal evidence of the degradation is unquestionable (e.g., noise due to the carrier degradation), without going beyond the technological level of that time.
- Aesthetical approach: it pursues a sound quality that matches the actual user's expectations (for both new commercial editions and to arrange the signal before the use of MIR techniques).
- Sociological approach: it has the purpose of obtaining a historical reconstruction of the recording as it was listened to at the time (see Storm, Type I in Sect. 3).
- Reconstructive approach: it has the objective to preserve the intention of the author (see Storm, Type II in Sect. 3).

In order to achieve one or more of these aims, it is necessary to have at disposal several audio restoration instruments. The audio restoration algorithms can be divided into three categories [15]:

1. frequency-domain methods, such as various forms of non-casual Wiener filtering or spectral subtraction schemes and recent algorithms that attempt to incorporate knowledge of the human auditory system; these methods use little a priori information;
2. time-domain restoration by signal models such as extended Kalman filtering (EKF): in these methods a lot of a priori information is required in order to estimate the statistical description of the audio events;
3. restoration by source models [16]: only a priori information is used.

The advantage of frequency-domain methods is that they are straightforward and easy to implement. However, the limitations are as follows: musical noise (short sinusoids randomly distributed over time and frequency) is unavoidable; the results depend on a good noise estimation. Restoration by source model is limited to very few cases (e.g., only monophonic recordings) and it is not generalizable. The EKF is able, in principle, to simultaneously solve the problems of filtering, parameter tracking and elimination of the outliers, but it is very sensitive to parameter setting (i.e., the order p of the AR model; the length q of the signal vector, the length of the initial training segment in the bootstrap procedure, the adaption speed λ, the forgetting factor γ, the threshold μ for detection of impulsive noise).

The author developed innovative algorithms, using the VST plug-in architecture, able to offer satisfying solutions to the problems, connected with the multiple carrier, peculiar to ethnic music recordings. The algorithms are detailed in the next subsections:

- Canazza REstoration Audio-extended Kalman filter (CREAK): A de-noise and de-click system based on EKF, dedicated to the restoration of audio signal re-recorded from shellac discs: SNR, clicks, pops, crackle.
- Canazza-Mian Suppression Rule (CMSR): A de-noise algorithm based on STSA (Short Time Spectral Attenuation), dedicated to the restoration of audio signal re-recorded from wax and amberol cylinders and shellac discs: low SNR.
- Perceptual audio restoration (PAR): A de-hiss based on perceptual algorithm for reel-to-reel tapes and cassettes: high SNR.

Of course, regardless their dedications, in a real restoration work it is opportune to combine these tools in order to obtain the better results.

4.1 CREAK: a de-noise and de-click system dedicated to shellac discs

In this tool, we employ an algorithm whose objective is to simultaneously solve the problems of filtering/parameter tracking/elimination of the outliers ("clicks") by using the EKF theory, as proposed by Niedzwiecki and Cisowski [17–19]. In particular the algorithm in [19] can be interpreted as the nonlinear combination of two Kalman filters: the first is used to follow the slow variations of the signal time-varying AR model parameters, while the second takes part in the reduction of background and impulsive noise. Because the old analogue discs (in particular: shellac discs) are corrupted by a broadband noise and by a large amount of impulsive disturbances (pops, clicks, and crackle), this algorithm is suitable for these carriers. In order to achieve maximum performance from the EKF, it is essential to optimize its implementation. For this purpose, to cope with the non-stationary nature of the audio signal, we used two properly combined EKF filters (forward and backward), and introduced a bootstrapping procedure for model tracking. The careful combination of the proposed techniques and an accurate choice of some critical parameters, allows to improve the performance of the EKF algorithm. The problem statement as well as some improvements on the filter stability and the tracker part are detailed in [20]. Here, we describe some improvements suitable for the restoration of shellac discs.

4.1.1 Bootstrap procedure

The first problem we deal with is the choice of the filter initial conditions. To this purpose, let us notice first that starting the algorithm from scratch implies an initial transient of the parameter tracker during which the EKF noise reduction capabilities are greatly reduced. To solve the problem, I have found useful to introduce a bootstrap procedure: the first 100 ms of the signal are time-reversed and fed to the filter. This way, parameters for a proper initialization of the model are estimated and restoration of the "true" signal will use these values as initial conditions.

4.1.2 Forward/backward filtering

The non-stationarity of the audio signal has an important consequence: the results of the forward and backward (reversing the time axes) filtering can be different. The algorithm is directional for its nature, that is, it uses the whole past history plus a finite number of future samples, depending on the model order. A provision that improves the algorithm performance is given by the use of two properly combined EKFs operating forward and backward on the signal [21].

It is clear that, with broadband noise, sharp changes in dynamics of the music signal are treated in a more effective way if they are "covered downhill" (i.e., passing from loud to soft intensity), independently from the direction of the filter. This is due to the fact that the estimate of the EKF benefits from having a signal segment with a better local Signal to Noise Ratio, before the transition loud/soft.

The comparison between the residuals of the forward and backward filtering shows that the former works better than the latter at the end of the restored segment, and that the opposite situation holds in the initial zone.

Furthermore, the forward/backward strategy improves the detection of impulsive disturbances: indeed, it can happen that the clicks are identified (and removed) in a more effective way in one direction than in the other [22].

Since the two filters give different signal estimates, $\hat{s}_+(t)$ (forward) and $\hat{s}_-(t)$ (backward), we found it effective to combine them according to:

$$\hat{s}_w(t) = \frac{\hat{\sigma}^2_{\varepsilon-}(t)}{\hat{\sigma}^2_{\varepsilon+}(t) + \hat{\sigma}^2_{\varepsilon-}(t)}\hat{s}_+(t) + \frac{\hat{\sigma}^2_{\varepsilon+}(t)}{\hat{\sigma}^2_{\varepsilon+}(t) + \hat{\sigma}^2_{\varepsilon-}(t)}\hat{s}_-(t) \tag{1}$$

The basic idea is to weigh $\hat{s}_+(t)$ and $\hat{s}_-(t)$ in a way that is inversely proportional to signal variance $\hat{\sigma}^2_{\varepsilon}$.

With such a provision it is possible, in the author's experience, to effectively remove broadband noise in audio signal with low SNR and, since we have two different click detectors, the effectiveness in removing impulsive disturbances is also improved. In this sense, it is particularly well-suited for the restoration of analogue discs.

4.2 CMSR: a de-noise algorithm dedicated to wax and amberol cylinders and shellac discs

The most widespread techniques (Short Time Spectral Attenuation, STSA) employ a signal analysis through the Short-Time Fourier Transform (which is calculated on small partially overlapped portions of the signal) and can be considered as a non-stationary adaptation of the Wiener filter in the frequency domain. The time-varying attenuation applied to each channel is calculated through a determined *suppression rule*, which has the purpose of producing an estimate (for each channel) of the noise power. A typical suppression rule is based on the Wiener filter [22]: usually the mistake made by this procedure in retrieving the original sound spectrum has an audible effect, since the difference between the spectral densities can give a negative result at some frequencies. Should we decide to arbitrarily force the negative results to zero, in the final signal there will be a disturbance, constituted of numerous random frequency pseudo-sinusoids, which start and finish in a rapid succession, generating what in literature is known as *musical noise*.

More elaborated suppression rules depend on both the relative signal and on a priori knowledge of the corrupted signal, that is to say, on a priori knowledge of the probability distribution of the under-band signals [22]. A substantial progress was made with the solution carried out in Ephraim and Malah [23], that aims at minimizing the mean square error (MSE) in the estimation of the spectral components (Fourier coefficients) of the musical signal. The gain applied by the filter to each spectral component does not depend on the simple Signal to Noise Ratio (Wiener Filter), but it is in relation with the two parameters Y_{prio} (SNR calculated taking into account the information of the preceding frame) and Y_{post} (SNR calculated taking into account the information of the

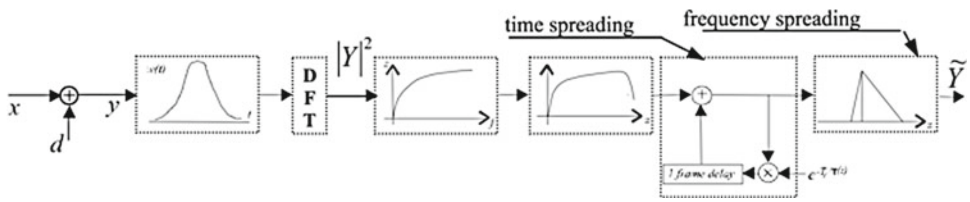

Fig. 1 The audio signal transformation from "outer to "inner representation. The signal $x(n)$ is first windowed by the $w(n)$ window and transformed in the frequency domain. The short time spectral power is transformed from Hertz (f) to Bark (z) scale, band-limited and spread both in time and frequency

current frame). A parameter (α) controls the balance between the current frame information and that of the preceding one. By varying this parameter, the filter smoothing effect can be regulated. Y_{prio} has less variance than Y_{post}: this way, musical noise is less likely to occur (see [23,24] for details).

Unfortunately, in the case of cylinders or shellac discs an optimal value of α does not exist, as it should be time-varying (because of the cycle-stationary characteristics of the cylinder/disc surface corruptions). Considering this, the author has developed a new suppression rule (Canazza-Mian[4] Suppression Rule, CMSR), based on the idea of using a *punctual* suppression without memory (Wiener like) in the case of a null estimate of Y_{post}, according to:

$$\alpha = \begin{cases} 0.98, & \text{if } Y_{\text{post}}(k, p) > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

The experiments carried out confirm that the filter performs very well, with a noise removal decidedly better than other suppression rules (e.g., classic EMSR) and with the advantage of not introducing musical noise, at least for SNR $\in [0 \div 20]$ dB (a typical value in the audio signal re-recorded from the cylinder and shellac discs). Furthermore, the behavior in the transients is similar of the EMSR filter, without having the perceptual impression of a processing "low-pass filter" like.

4.3 PAR: a de-hiss perceptual algorithm dedicated to reel-to-reel tapes and cassettes

This tool considers the perceptually relevant characteristics of the signal. This way, within model fidelity, only the audible noise components are removed in order to preserve the signal from possible distortions caused by the restoration process. In this sense, this method is particular suitable for the restoration of signals with a high SNR (SNR > 20 dB).

[4] Mian (1942–2006) was a professor of Digital Signal Processing at the Dept. of Information Engineering, University of Padua, a leading researcher in our department, and an outstanding teacher whose brightness and kindness I will always remember. These results are affectionately dedicated to his memory.

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 [25], sketched in Fig. 1. 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, band-limited 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 $\tilde{Y}(p, z) \approx \tilde{X}(p, z) + \tilde{D}(p, z)$, defined in the Bark domain, band-limited and processed taking into account the spreading both in time and frequency. Finally, the $\tilde{Y}_{\text{prio}}$ and $\tilde{Y}_{\text{post}}$ terms (see Sect. 4.2) are calculated according to the inner representation and the gain $\tilde{G}(p, z)$ is derived. Figure 2 shows a representative example: a sinusoid with broadband noise (top) and after the perceptual de-noise (bottom), in which it can be noticed that only the audible noise components are removed.

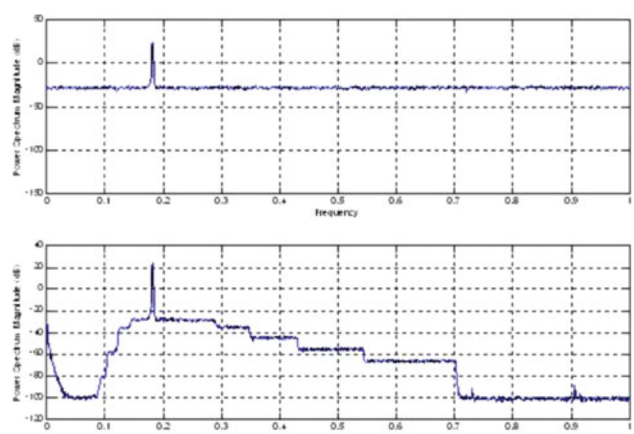

Fig. 2 A sinusoid with broadband noise (*top*) and after the perceptual de-noise (*bottom*). Only the audible noise components are removed. X-axis: frequency normalized to the Nyquist frequency; Y-axis: power spectrum magnitude (dB)

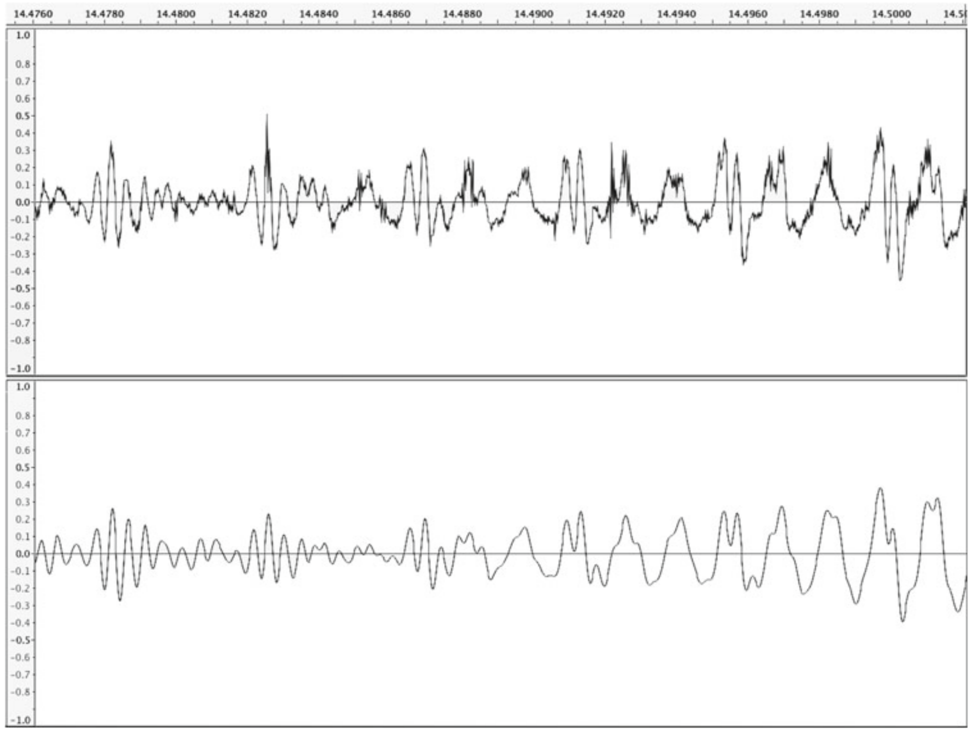

Fig. 3 *Top* The waveform of the original (corrupted) audio extract. *Bottom* The reconstructed data by CMSR. The increase of SNR can be noticed. *X*-axis: time (s). *Y* axis: amplitude (normalized)

5 Experimental results

We conducted a series of experiments with real usage data from different international audio archives. In this section we present our experimental results of applying the above described techniques related to audio restoration. As first case study, Fig. 3 shows a restoration of a wax cylinder by means of CMSR (see Sect. 4.2). The song is *My Mariuccia take-a steamboat*, performed by Billy Murray (vocal tenor) in 1906. Edison Gold Moulded Record: 9430; cylinder length: 2' 13". It is a comic song in Italian dialect with orchestra accompaniment. In Fig. 3: at the top there is the waveform of the original (corrupted) audio extract, at bottom, the restored data by means of CMSR. Only a de-noise is performed. An increase of SNR can be noticed.

Considering impulsive disturbances, the Fig. 4 shows a de-click of a shellac disc by means of CREAK (see Sect. 4.1). The song is *La signorina sfinciusa* (The funny girl), performed by Leonardo Dia. Shellac 78 rpm 10", Victor V-12067-A (BVE 53944-2); disc length: 3' 19". The lyrics are in an Italian dialect, with the musical accompaniment of a mandolin (Alfredo Cibelli) and two guitars (unknown players). Recorded in New York, July, 24, 1929. In Fig. 4 is pointed out a click, before (top) and after (bottom) the audio restoration performed by CREAK.

In a third case study, a tape recording (unpublished) of Portuguese fado music (from the audio archive of the Universidade Nova de Lisboa, Faculdade de Ciencias Sociais e Humanas, Portugal[5]) is considered. In this case we performed only a de-noise by means of PAR tool. Figures 5 and 6 show the corrupted (top) and restored (bottom) signals respectively in time and frequency domains of two different (representative) excerpts of the musical piece.

Finally, an example of a combined methods is presented. We consider the shellac disc *Nofrio e la finta americana*, performed by Giovanni De Rosalia and Francesca Gaudio (vocals). Shellac 78 rpm 10", Victor 72404 B (B 22911-2); disc length 2' 40". Recorded in New York, June 11, 1919. In this case, we carried out de-click and de-noise by means of CREAK. Because of the low SNR (SNR $\sim$ 5 dB), we processed the signal also with CMSR. In this way, we obtained a SNR = 40 dB[6], without introducing particularly audible distortions (musical noise). Figures 7 and 8 show the corrupted (top) and restored (middle and bottom) signals respectively in the time and frequency domains.

5.1 Comparison

Figure 9 shows the gain trend introduced by the filters described above in comparison with some *standard* filters (Wiener filter, Power Subtraction [22], EMSR [23,24]) at

[5] The author would like to thank Salwa Castelo Branco for sharing the audio documents of the archive.

[6] The measuring of noise power is made by taking the noise print in an interval where there is only background noise.

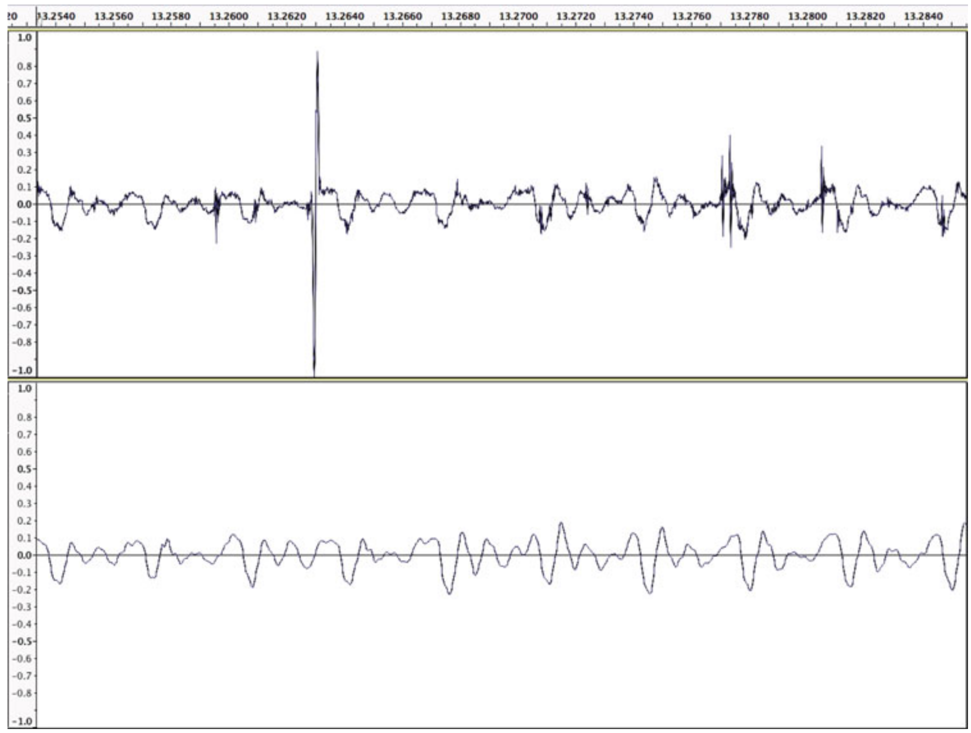

Fig. 4 *Top* The waveform of the original (corrupted) audio extract. *Bottom* The reconstructed data by CREAK. The click removal can be noticed. *X*-axis: time (s). *Y*-axis: amplitude (normalized)

the varying of the noisy signal SNR, considering 35 carriers of ethnic music (20 shellac discs recorded from 1910 to 1930, 6 wax cylinders recorded from 1900 and 1914, and 9 open-reel music tapes recorded from 1960 to 1975). The term gain indicates the difference between the de-noised signal SNR and the input signal SNR. As can be noticed, all the three filters have a good performance, in particular CMSR for signal with low or medium SNR, CREAK for SNR [15 ÷ 30] dB and PAR seems adapt to reduce the noise in audio signal with a good SNR.

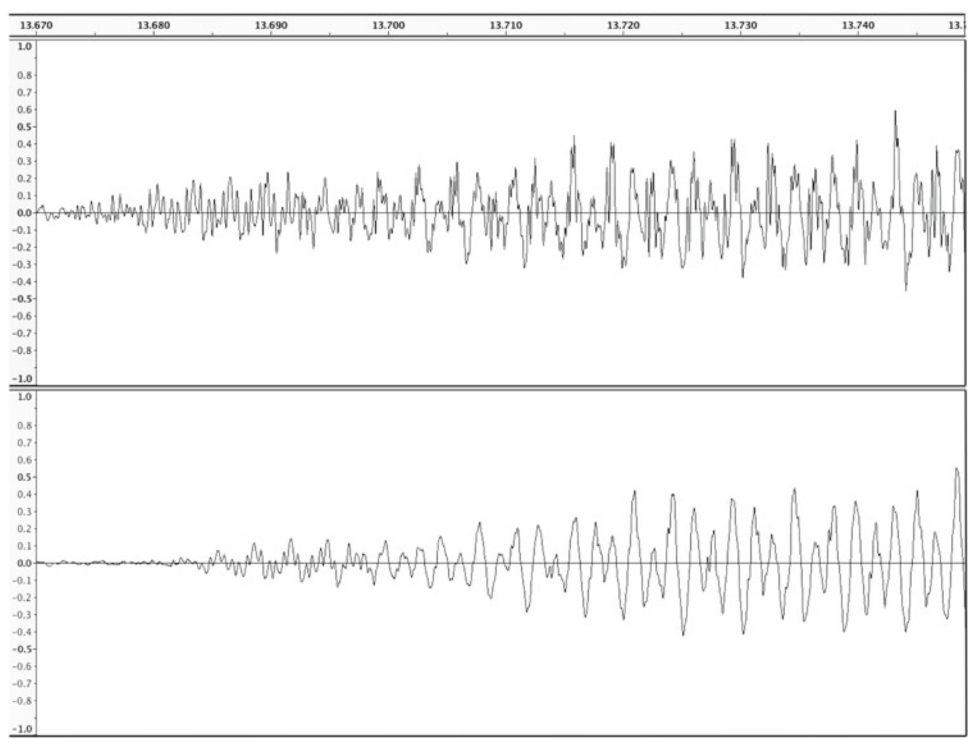

Fig. 5 *Top* the waveform of the original (corrupted) audio extract. *Bottom* the restored data by PAR. *X*-axis: time (s). *Y*-axis: amplitude (normalized)

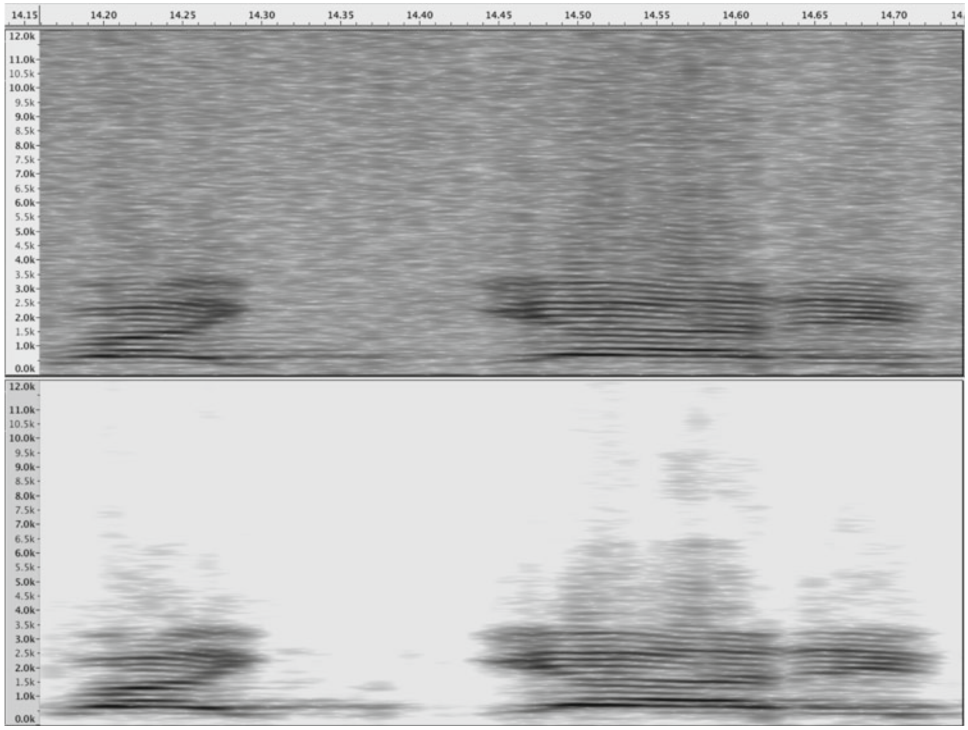

Fig. 6 *Top* the spectrum of the original (corrupted) audio extract. *Bottom* the restored data by PAR. X-axis: time (s). Y-axis: frequency (Hz)

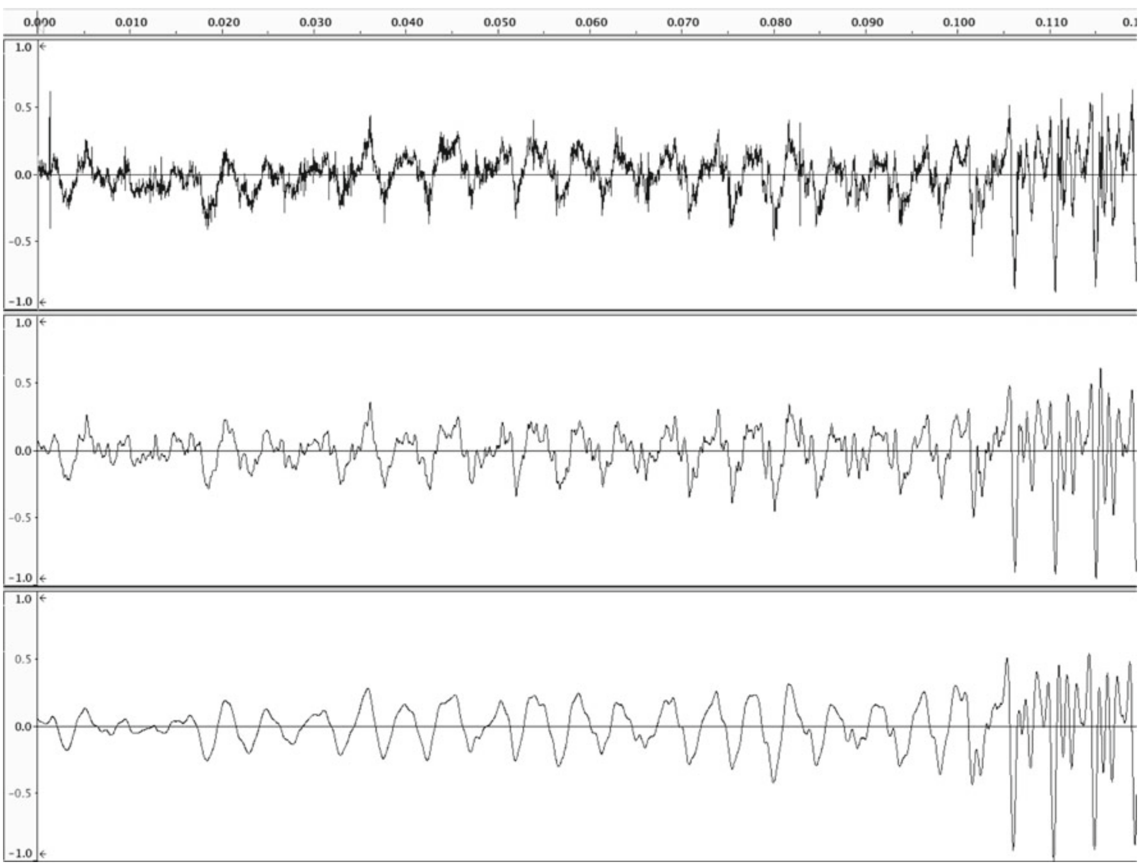

Fig. 7 *Top* the waveform of the original (corrupted) audio extract. *Middle* de-clicked and de-noised by CREAK. *Bottom* de-noised by CMSR. X-axis: time (s). Y-axis: amplitude (normalized)

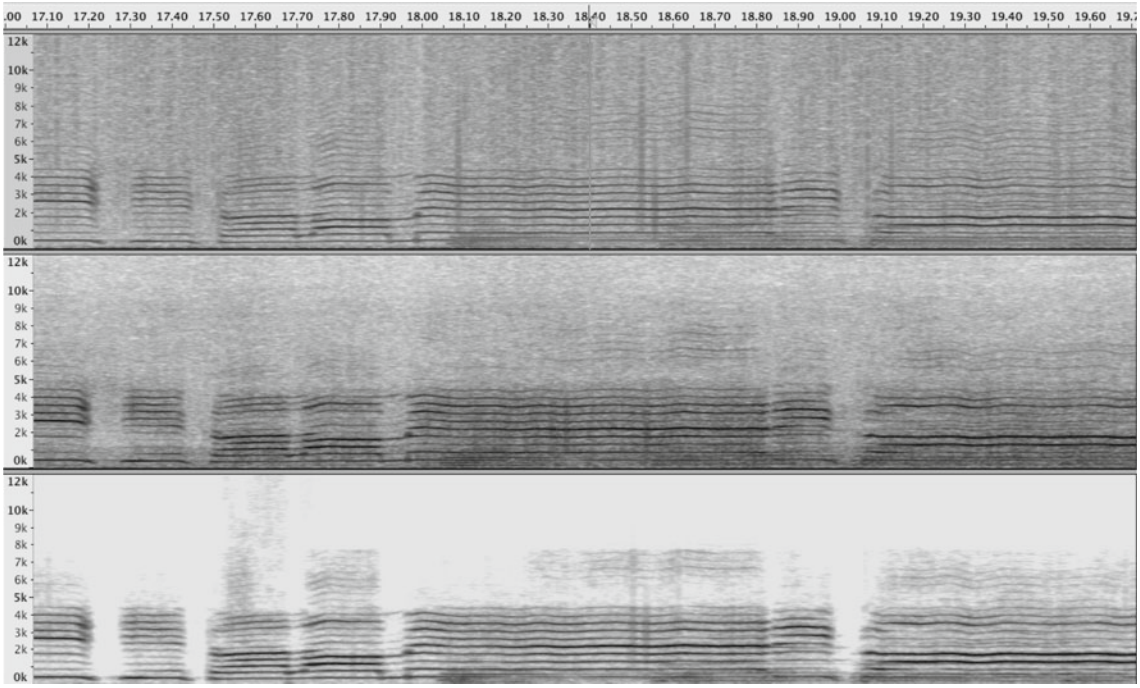

Fig. 8 *Top* the spectrum of the original (corrupted) audio extract. *Middle*: de-clicked and de-noised by CREAK. *Bottom* de-noised by CMSR. X-axis: time (s). Y-axis: frequency (Hz)

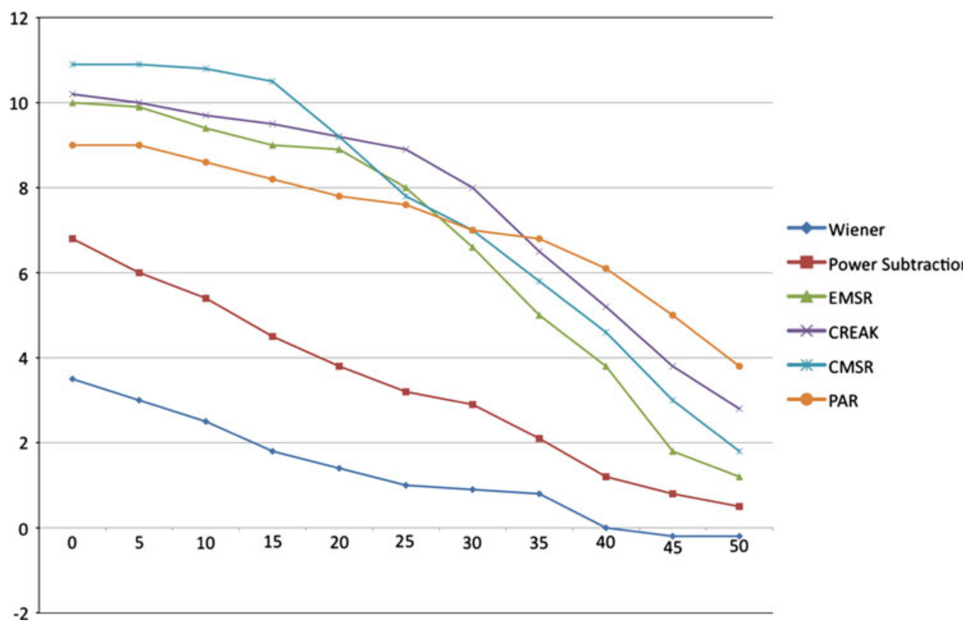

Fig. 9 Gain trend introduced by the filters in the frequency domain at the varying of the input SNR (SNR_{out}-SNR_{in} vs. SNR_{in} in dB). The three filters have a good performance: CMSR for signal with low or medium SNR; CREAK for SNR [15 ÷ 30] dB; PAR for high SNR

6 Assessment

To validate the system, a listening test was carried out.

6.1 Material

Three music pieces recorded in clearly different supports were used (see Sect. 5 for the details of these audio documents):

(1) Cylinder: *My Mariuccia take-a steamboat*.
(2) Shellac disc: *La signorina sfinciusa*.
(3) Tape: unpublished recording of Portuguese fado music.

In order to minimize the fatigue and maximize the attention of the participants, the first 20 s of each track were selected. Since the task was more a comparison than an individual analysis, those short extracts seemed to be sufficient.

Restoration of the noisy stimuli was performed by CREAK in de-click mode, then using CREAK, CMSR and PAR as well as through the following commercial products:

(1) X-Click and X-Noise of Waves Restoration bundle (Waves V6 Update 2);
(2) De-click and De-noise of Sonic NoNOISE plug-in suite (a software tool ables to offer the same algorithms found in the SonicStudio version);
(3) Declicker and Denoiser (enable its *Musical noise suppression* filter) of iZotope RX v1.06;
(4) Auto Declick and Auto Dehiss of CEDAR Tools;
(5) Adobe Audition 3.0;
(6) Audacity 1.3.6 (an open source software for recording and editing audio signals).

The plug-ins by Sonic NoNOISE and by CEDAR are used in a Pro Tools HD 2 system. The parameters used to control the different systems were subjectively set to obtain the best tradeoff between noise removal and music signal preservation. This way, 27 restored stimuli were produced.

6.2 Test method

The tests were conducted using the EBU MUSHRA test method [6], which is a recommended evaluation method adopted by ITU [26]. This protocol is based on "double-blind triple-stimulus with hidden reference" method, which is stable and permits accurate detection of small impairments. An important feature of this method is the inclusion of the hidden reference and two bandwidth-limited anchor signals (7 and 3.5 kHz).

6.3 Training phase

The purpose of the training phase, according to the MUSHRA specification, was to allow each listener: i) to become familiar with all the sound excerpts under test and their quality-level ranges; ii) to learn how to use the test equipment and the grading scale.

6.4 Listeners

Two subject groups were selected: 12 researchers of Universities of Padua and Udine and 14 students of the Multimedia Communication Degree (University of Udine). All the subjects were musically trained.

6.5 Equipment

The audio signals were recorded at 96 kHz/24 bit (uncompressed sound files) and played through Apple iMac Intel Core 2 Duo with 2 GB 800 MHz DDR2 SDRAM (D/A converter: RME Fireface 400), and headphones (AKG K 501). The listeners could play all the stimuli under test any order they liked, including the hidden reference and the two bandwidth-limited anchor signals.

6.6 Test duration

The training session for each listener took $\sim$ 1 h, including an explanation about the tests and equipment, and a practice grading session. The grading phase consisted of 3 test sessions (one for each music piece), each one containing 12 test signals (1 noisy signal, 9 restored signals, 2 anchors). Each session took, on average, about 10 min. Subjects were allowed a rest period between each session, but not during a session.

6.7 Main results

The statistical analysis method described in the MUSHRA specification was used to process the test data. The results are presented in Table 1 as mean grades. The results from two listeners were removed because the mean of their rates (in absolute value) on hidden references was greater than $+/-0.5$.

In *My Mariuccia take-a steamboat* CEDAR and CMSR are the only restoration system with a score > 3; the other software produced lower scores, with fair quality assessments. The anchor at 7 kHz obtains a score greater than 0 (similar to those of a few commercial products). The quality range between the best and the worst restoration system is 2.60. CMSR produces scores greater than NoNOISE; CREAK and noNOISE achieve the same score.

Table 1 Mean for restored stimuli and anchors, 24 subjects

Restoration system	S1	S2	S3	Grand average
CMSR	+3.10	+3.55	+2.80	3.15
CREAK	+2.90	+3.95	+3.00	3.28
PAR	+2.00	+1.20	+4.45	2.55
CEDAR Tools	+3.20	+3.70	+4.25	3.72
NoNOISE	+2.90	+1.78	+4.00	2.89
iZotope RX	+2.40	+1.70	+2.40	2.17
Waves	+2.30	+1.55	+3.45	2.43
Audacity	+0.60	+1.25	+3.20	1.68
Adobe Audition	+0.60	+0.55	+2.25	1.13
Anchor 7 kHz	+0.50	+0.00	−2.50	−0.67
Anchor 3.5 kHz	−1.00	−4.00	−5.00	−3.33

Stimuli: S1 = *My Mariuccia take-a steamboat*; S2 = *La signorina sfinciusa*; S3 = Unpublished tape recording

The best restoration systems for *La signorina sfinciusa* are CEDAR, CREAK and CMSR. The anchor at 7 kHz produces scores equal to 0. The quality range between the best and worst restoration system is only 3.45.

The best restoration systems for the stimuli recorded on tape are PAR and CEDAR; all the other software produced similar scores, with low quality assessments. The anchors produces scores below 0. The quality range between the best and worst restoration system is only 2.2.

6.8 Discussion

It is possible to make some important comments:

– Observing the quality range between the best and worst restoration system, it seems reasonable to conclude that all the restoration algorithms work quite well (i.e., the user's evaluation is good enough) with high SNR signals (SNR $>$ 30 dB) as well as with very low SNR stimuli (SNR $<$ 10 dB): see the scores achieved with the *My Mariuccia take-a steamboat* and the tape stimuli.
– the behavior is in connection with the results of the objective comparison carried out (see the results showed in Sec 5.1). Summarizing: CMSR for the low and medium SNR, CREAK for medium SNR and PAR for high SNR.
– The best single tool is CEDAR, with a grand average equal to 3.72 (see Tab. 1). However, let us consider the three tools CREAK, CMSR, and PAR as a single restoration environment (it could be called *CARE tool*: CAnazza REstoration or Csc Audio REstoration[7]): in this sense, the grand average of CARE is 3.83. This result explains the expedience to develop (and to use, of course) different tools, in relation to the supports considered.

7 Concluding remarks

The opening up of archives and libraries to a large telecoms community represents a fundamental impulse for cultural and didactic development. Guaranteeing an easy and ample dissemination of some of the fundamental moments of the musical culture of our times is an act of democracy which cannot be renounced and which must be assured to future generations, even through the creation of new instruments for the acquisition, preservation and transmission of information. This is a crucial point, which is nowadays the core of the reflection of the international archival community. If, on the one hand, scholars and the general public started paying greater attention to the recordings of artistic events, on the other hand, the systematic preservation and access to these documents is complicated by their diversified nature and amount.

Ethno-musicologists, scholars, audio archives personnel usually need to use a large—hardly manageable—number of sources stored in different media: outlines and annotations, scores, theater programmes and critical reviews, setting photos, audio signal and video footages. Although over the past few years the European Union has provided fundings for many projects focused on text codification and the creation of editions in electronic format, in most cases, modeling the traditional ecdotics methods, these studies apply technology without a real sharing of models and methodologies already in use by the information science [4]. In the last few years, the great success of the Web has made the idea of using the hypertextual philosophy, on which it is based, attractive; so, many people have adopted, uncritically at times, ill-organized hypertextual structures with intrinsic problems of confusion and cognitive overload: the models, which have been used by human science so far, do not allow the user to separate structure and contents, or to compare different media at various abstraction levels, or to use the system in a personalized and adaptive way, or to aggregate the data for generating new knowledge. For this purpose, the author proposed a new model, called PSYCHO-MAD, a Powerful SYstem with Charming Hypermedia Objects for Music Audio Documents. This is a not-hierarchical hypermedia model for handling the information stored in the audio memories based on an extension of zz-structures. The cooperation activities of different classes of actors allow the user to create new virtual hyperdocuments and dynamic views (useful, for example, in ethno music events). The model connects, without preconceived limitations, documents stored in different media: annotations made by the author, scores, room programs, critical reviews, setting photos, sound recordings and video shootings. PSYCHO-MAD aims to organize and manage data and knowledge using a distributed, agent-based extension of innovative data structures, the zz-structures [27,28] and it is conceptually based on:

– the idea of the musical work as an open system, that is, a multidimensional object made up of the text, event, tradition, interpretation, fruition and reception;
– the hypertextual perspective: the used hypermedia model connects, without any previously formed restrictions, the contextual information recreated in the musical work from an historic-musicological, biographical and philological points of view and allows the user to listen different texts and interpretations of the same musical work and to compare them to scores (and other direct or indirect sources), musicological analyses and audiovisual recordings.

[7] CSC stands for Centro Sonologia Computazionale, the international laboratory of the University of Padua, leader in the Sound and Music Computing field since 1979.

Please see [4] for a detailed description of the model and some case studies related to ethnic music.

The main goal of this paper is to stress that the archiving process of digitized audio documents is complete only when it includes all the ancillary information, in particular the contextual information of the original carriers. In this sense guidelines to the A/D transfer are detailed in Sect. 3, in order to minimize the information loss and to automatically measure the unintentional alterations introduced by the A/D equipment. The author strongly believes that the restoration process must start from the analysis of the ancillary information, in order not to remove intentional signal. Moreover, it is particularly important to have at disposal several restoration tools, customized to different kind of audio recordings (in the same way that audio engineers use different microphones for different musical instruments). This work has presented:

– The CARE tool (Sect. 4). The ethnic music audio documents are usually recorded in non-professional carriers by means of amateur recording system. Thus, for their appropriate fruition and/or for a suitable use of MIR techniques is necessary to process the audio signals by means of audio restoration techniques.
– Four different case studies (Sect. 5), carried out using different carriers.
– An objective comparison, in order to validate the system, of the CARE tool with some *standard* filters at the varying of the SNR, considering 35 ethnic music documents (Sect. 5.1).
– A perceptual assessment, using the EBU MUSHRA test protocol [6] (Sect. 6).

The extensive tests, carried out to investigate the effectiveness of the suggested algorithms when applied to a variety of audio data, show that the tool presented outperforms standard approaches to restoration. Listening tests (Sect. 6) confirm practical usefulness of the proposed solutions. The algorithm are implemented as a plug-in based software tool, which can be used as an added module to the most commonly used audio editors.

This work summarizes a number of experiences in several research/applied project on ethnic music audio archives, carried out by the author, including: “Electronic Storage and Preservation of Artistic and Documentary Audio Heritage (speech and music)” funded by the National Research Council of Italy (CNR); “Preservation and Online Fruition of the Audio Documents from the European Archives of Ethnic Music” funded by the EU under the Program Culture2000; “Preservation and Access of the Fugazzotto Archive: the songs of Italian immigrants”; Archives of popular tradition: preservative re-recording and cataloguing of popular music collection in “V. Joppi” Civic Library of Udine.

References

1. Middleton, R.: Studying Popular Music. Open University Press, Philadelphia (1990)
2. Yakel, E.: Digital curation. OCLC Syst. Serv. **23**(4), 335–340 (2007)
3. Bernardini, N., Bresin, R., Camurri, A., De Poli, G., Leman, M., Rocchesso, D., Serra, X., Valimaki, V., Widmer, G.: A roadmap for sound and music computing. In: The S2S2 Consortium (2007)
4. Canazza, S., Dattolo, A.: The past through the future: a hypermedia model for handling the information stored in the audio memories. J. New Music Res. **38**(4), 381–396 (2009)
5. Zattra, L.: The assembling of Stria by John Chowning: a philological investigation philological investigation. Comput. Music J. **31**(3), 38–64 (2007)
6. ITU-R: Methods for the Subjective Assessment of Small Impairments in audio systems including multi-channel sound systems. Recommendation BS.1116-1 (2000)
7. Orio, N., Snidaro, L., Canazza, S., Foresti, G.L.: Methodologies and tools for audio digital archives. Int. J. Digit. Libr. **10**(4), 201–220 (2009)
8. Storm, W.: The establishment of international re-recording standards. Phonogr. Bull. **27**, 5–12 (1980)
9. Schüller, D.: Preserving the facts for the future: principles and practices for the transfer of analog audio documents into the digital domain. J. Audio Eng. Soc. **49**(7-8), 618–621 (2001)
10. Boston, G.: Safeguarding the Documentary Heritage. A guide to Standards, Recommended Practices and Reference Literature Related to the Preservation of Documents of all kinds. UNESCO, Paris (1988)
11. Brock-Nannestad, G.: The objective basis for the production of high quality transfers from pre-1925 sound recordings. In: AES Preprint No. 4610. Audio Engineering Society 103rd Convention, pp. 26–29. New York (1997)
12. AES-11id-2006: AES Information Document for Preservation of Audio Recordings—Extended Term Storage Environment for Multiple Media Archives. AES (2006)
13. IASA-TC 03: The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy. IASA Technical Committee (2005)
14. Orio, N., Snidaro, L., Canazza, S.: Semi-automatic metadata extraction from shellac and vinyl disc. In: Proceedings of Workshop on Digital Preservation Weaving Factory for Analogue Audio Collections (2008)
15. Canazza, S., Vidolin, A.: Preserving electroacoustic music. J. New Music Res. **30**(4), 351–363 (2001)
16. Esquef, P.A.A., Valimaki, V., Karjalainen, M.: Restoration and enhancement of solo guitar recordings based on sound source modeling. J. Audio Eng. Soc. **50**(4), 227–236 (2002)
17. Niedźwiecki, M.: Steady-state and parameter tracking proprieties of self-tuning minimun variance regulators. Automatica **25**, 597–602 (1989)
18. Niedźwiecki, M.: Identification of time-varying processes in the presence of measurement noise and outliers. In: Proceedings of 11th IFAC Symposium on System Identification, pp. 1765–1768 (1997)
19. Niedźwiecki, M., Cisowski, K.: Adaptive scheme for elimination of broadband noise and impulsive disturbances from AR and ARMA signals. IEEE Trans. Signal Process **44**(3), 967–982 (1996)
20. Canazza, S., De Poli, G., Mian, G.A.: Restoration of audio documents by means of extended Kalman filter. IEEE Trans. Audio Speech Lang. Process **18**(6), 1107–1115 (2010)
21. Anderson, B.D.O., Moore, J.B.: Optimal filtering. Prentice-Hall, Englewood Cliffs, NJ (1979)

22. Godsill, S., Rayner, P., Cappé, O.: Digital audio restoration. Kluwer, Boston, MA (1998)
23. Ephraim, Y., Malah, D.: Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process **32**(6), 1109–1121 (1984)
24. Ephraim, Y., Malah, D.: Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Trans. Acoust. Speech Signal Process **33**(2), 443–445 (1985)
25. Beerends, J.G., Stemerdink, J.A.: A perceptual audio quality measure based on psychoacoustic sound representation. J. Audio Eng. Soc. **40**(12), 963–978 (1992)
26. EBU Project Group B/AIM: EBU report on the subjective listening tests of some commercial internet audio codecs (2000)
27. Nelson, T.: A cosmology for a different computer universe: data model, mechanisms, virtual machine and visualization infrastructure. J. Digit. Inf. **5**(1), 311–319 (2004)
28. Dattolo, A., Luccio, F.: A state of art survey of zz-structures. In: Proceedings of the New Forms of Xanalogical Storage and Function Workshop. 20th ACM Conference on Hypertext and Hypermedia, Hypertext 2009, pp. 1–6. Torino, Italy (2009)