Chapter 11

Musical cultural heritage: From preservation to restoration

Sergio Canazza

Copyright © 2005-2013 Sergio Canazza
except for paragraphs labeled as adapted from <reference>
This book is licensed under the CreativeCommons Attribution-NonCommercial-ShareAlike 3.0 license. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/, or send a letter to Creative Commons, 171 2nd Street, Suite 300, San Francisco, California, 94105, USA.

11.1 Introduction

The availability of digital archives and libraries on the Web represents a fundamental impulse for cultural and didactic development. Guaranteeing an easy and ample dissemination of some of the fundamental moments of the music culture of our times is an act of democracy that must be assured to future generations, even through the creation of new tools for the acquisition, preservation, and transmission of information. This is a crucial point, which is nowadays one of the core reflections of the international archive community. If, on the one hand, scholars and the general public have begun paying greater attention to the recordings of artistic events, on the other hand, the systematic preservation and consultation of these documents is complicated by their diversified nature, because the data contained in the recordings offer a multitude of information on their artistic and cultural life, that goes beyond the audio signal itself.

In this sense, a complete access to the audio content cannot be carried out without accessing to the contextual information, that is to all the content-independent information available from the cover, the signs on the carrier, and so on. In addition, a preservative re-recording and cataloguing of audio document collections cannot leave out a consideration of the history of the institutions or collections in which they are held. In fact, this information helps defining the strategy to adopt during the preservative interventions.

It is well-known that the recording of an event can never be a neutral operation, because the timbre quality and the plastic value of the recorded sound, which are of great importance in contemporary music, are already influenced by the positioning of the microphones used during the recording. In ad-
dition, the audio processing carried out by the Tonmeister\textsuperscript{1} is a real interpretative element added to the recording of the event. Thus, musicological and historic-critical competence becomes essential for the individuation and correct cataloguing of the information contained in audio documents. Being made of unstable base materials, sound carriers are more subject to damage caused by inadequate handling. The commingling of a technical and scientific formation with historic-philological knowledge becomes essential for preservative re-recording operations, going beyond mere analog-to-digital (A/D) transfer.

Since the first recording\textsuperscript{2} on paper made in 1860 (by Edouard-Léon Scott de Martinville “Au Clair de la Lune” using his phonautograph) to the modern Blu-ray Disc, what we have in the audio carriers field today is a Tower of Babel: a bunch of incompatible analogue and digital approaches and carriers – paper, wire, wax cylinder, shellac disc, film, magnetic tape, vinyl record, magnetic and optical disc, to mention only the principal ones – without standard players able to read all of them. As far as audio memories are concerned, preservation is divided into a passive\textsuperscript{3} preservation, that is the defence of the carrier from external agents without altering the structure, and an active preservation, which involves data transfer on new media. 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!”. 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”. 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.

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

\textsuperscript{1}The term Tonmeister describes a person who has a detailed theoretical and practical knowledge of all aspects of sound recording. But, unlike a sound engineer, he/she must be also deeply musically trained. Both competencies have equal importance in a Tonmeister’s work.

\textsuperscript{2}Unlike Edison’s similar 1877 invention, the phonograph, the phonautograph only created visual images of the sound playback capabilities. Scott de Martinville’s device was used only for scientific investigations of sound waves.

\textsuperscript{3}Passive preservation is divided into indirect, which does not physically involve the carrier, and direct, in which the carrier is treated without altering its structure and composition.
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.

It is worth noting that, in the Seventies/Eighties of 20th Century, expert associations (Audio Engineering Society: AES; National Archives and Records Administration: NARA; Association for Recorded Sound Collections: ARSC) were still concerned about the use of digital recording technology and digital storage media for long-term preservation. They recommended re-recording of endangered materials on analogue magnetic tapes, because of: a) rapid change and improvement of the technology, and thus rapid obsolescence of hardware, digital format and storage media; b) lack of consensus regarding sample rate, bit depth and record format for sound archiving; c) questionable stability and durability of the storage media. The digitization was considered primarily a method of providing access to rare, endangered, or distant materials – not a permanent solution for preservation. Abby Smith (director of programs at the Council on Library and Information Resources (CLIR), USA, http://www.clir.org), still in 1999, suggested that digitization should be considered a means for access, not preservation – “at least not yet”.

Nowadays, it is well-known that preserving the carriers and maintaining the dedicated equipment for their reproduction is hopeless. The audio information stored in obsolete formats and carriers is in risk of disappearing. To this end, the audio preservation community introduced the concept “preserve the content, not the carrier”. Audio (and video) preservation must therefore be based on digital copying of contents. Consequently, analogue holdings must be digitized. At the end of the 20th century, the traditional “preserve the original” paradigm shifted to the “distribution is preservation” idea of digitizing the audio content and making it available using digital libraries technology. Now the importance of transferring into the digital domain (active preservation) is clear, namely for carriers in risk of disappearing, respecting the indications of the international archive community (e.g., Audio Engineering Society, AES; International Association of Sound and Audiovisual Archives, IASA; International Federation of Library Associations, IFLA).

This chapter, after a detailed overview of the debate evolved since the Seventies inside the archivist community on audio documents preservation (Sect. 11.2), describes the protocols defined, the processes undertaken, the results ascertained from several international audio documents preservation projects and the techniques used. In particular, in Sect. 11.3 and Sect. 11.4, some guidelines are given, including recommendations to the A/D process directed to minimize the information loss and to automatically measure the unintentional alterations introduced by the A/D equipment, focusing on the high quality/high cost/low throughput cases. The author believes that the increased dimensionality of the data contained within an audio digital library should be dealt with by means of automatic annotations. Therefore, this chapter presents in Sect. 11.5 a set of tools able to extract, in a semi-automatic way, metadata from photos and video shootings of audio carriers. these tools are useful, in particular, in settings where it is necessary to put attention to the cost-benefit tradeoffs. Sect. 11.6 presents an original system for reconstructing the audio signal from a still image of a disc surface and an alignment technique aimed at comparing the effectiveness and the robustness of different re-recording techniques.

### 11.2 Audio Documents Preservation

A reconnaissance on the most significant positions of the debate evolved since the Seventies inside the archivist community on the audio documents active conservation highlights at least three different points of view, described below.
11.2.1 “Two Legitimate Directions”

It was William Storm, at that time Assistant Director of the Thomas A. Edison Re-recording Laboratory Syracuse University Libraries, who focussed on the problem of standardizing the procedures of audio restoration in an article which became famous for the numerous controversies it arose. Storm individuated two legitimate directions, 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. Storm’s contribution aimed at shifting the archivist’s interest from the simple collecting of audio carriers to the information contained in the recording, and at highlighting the double documentary value of re-recording by proposing an audio-history sound preservation: on the one hand, he wanted to offer a historically faithful reproduction of the original audio recording by extracting the sound content according to the historical conditions and technology of the era in which it was produced; on the other hand, he wanted to document the quality of sound reception offered by the recording and reproducing systems of the time. These two instances, conceptually joined in a single type of re-recording, had induced Storm to prescribe the use of original playback equipment. The aim of history preservation is to first hear how records originally sounded to the general public.

The second type of re-recording (Type II) was presented by Storm as a further stage of audio restoration, as a more ambitious research objective, conceived as a coherent development of Type I: The knowledge acquired through audio-history preservation provides the sound engineer with a logical place to begin the next step – the search for the true sound of an artist. Type II is then characterized by the use of playback equipment other than that originally intended so long as the researcher proves that the process is objective, valid, and verifiable, with the intent of obtaining the live sound of original performers, transcending the limits of a historically faithful reproduction of the recording.

11.2.2 “To Save History, Not Rewrite It”

The “Safeguarding the Documentary Heritage. A Guide to Standards, Recommended Practices and Reference Literature Related to the Preservation of Documents of All Kinds” commissioned by UNESCO reports the philosophical approach save history, not rewrite it. The audio section is clearly influenced by the new formulations made by Dietrich Schüller. Schüller’s works move from a different methodological point of view, which is to analyse what the original carrier represents, technically and artistically, and to start from that analysis in defining what the various aims of re-recording may be. Regarding the reconstruction of the history of music perception Schüller states: “The only case where the use of original equipment is justified is in the exotic aim to reconstruct the sound of a historical recording as it was heard originally”. Instead he points directly towards defining a procedure which guarantees the re-recording of the signal’s best quality by limiting the audio processing to the minimum. Having set aside the general philosophical themes, Schüller goes on to an accurate investigation of signal alterations which he classifies in two categories: intentional and unintentional. The former include recording, equalization, and noise reduction systems, while the latter are further divided into two groups: the ones caused by the imperfection of the recording technique of the time, resulting in various distortions and the ones caused by misalignment of the recording equipment, for example, wrong speed, deviation from the vertical cutting angle in cylinders or misalignment of the recording in magnetic tape.

The choice whether or not to compensate for these alterations reveals different re-recording strategies: historical faithfulness can refer to various levels: Type A the recording as it was heard in its time, which is equivalent to Storm’s Type I presented in the previous section; Type B the recording as it has been produced, precisely equalized for intentional recording equalizations, compensated for eventual er-
rors caused by misaligned recording equipment and replayed on modern equipment to minimize replay distortions.

Type B re-recording defines a historically faithful level of reproduction that, from a strictly preservative point of view, is preliminary to any further possible processing of the signal. These compensations use knowledge which is external to the audio signal; therefore, even in the operations provided for by Type B, there is a certain margin of interpretation because a historical acquaintance with the document is called into question alongside with technical-scientific knowledge. For instance, to individuate the equalization curves of magnetic tapes or to determine the rotation speed of a record. Most of the information provided by Type B is retrievable from the history of audio technology, while other information is instead experimentally inferable with a certain degree of precision. The re-recording work can thus be carried out with a good degree of objectivity and represents an optimal level within which the standard for a preservation copy can be defined.

After having established an operational criterion for preservative re-recordings, based on stable procedures and derived from an objective knowledge of the degradations, Schuller individuated a third level of historically faithful reproduction, type C: The recording as produced, but with additional compensation for recording imperfections caused by the recording technique of the time. While the compensations of type B are commonly accepted and must – as Schuller writes – be carried out, in type C they have to do with the area of equalizations used to compensate for non-linear frequency response, caused by imperfect historical recording equipment and to eliminate rumble, needle noise, or tape hiss. These are operations which elude standard operational criteria and must therefore be rigorously documented by the restorer, who must write out accurate reports in which he specifies both the equipment and systems used as well as all the restoration phases.

11.2.3 “Secondary Information”: the History of the Audio Document Transmission

The studies of George Brock-Nannestad are in line with the modeling of the degradations through reverse engineering. In these studies he focused on the A/D conversion of acoustic recordings (thus recordings made before 1925) and, in particular, the strong line spectrum in the recording transfer function and unknown recording speed. Brock-Nannestad goes back to the first studies in the acoustics of sound reproduction and to the scientific works of Dayton C. Miller, whom we must recall as the first to attempt to retrieve the true sound once it had been recorded. In order to be consistent and have scientific value, the re-recording work requires a complete integration between the historical-critical knowledge which is external to the signal and the objective knowledge which can be inferred by examining the carrier and the degradations highlighted by the analysis of the signal.

11.2.4 The Audio Preservation Protocol

Starting from these positions, I define a preservation copy a digital data set that groups the information carried by the audio document, considered as an artifact. It aims to preserve the documentary unity, and its bibliographic equivalent is the facsimile or the diplomatic copy. Signal processing techniques are allowed only when they are finalized to the carrier restoration. The audio format identification and the choice of the playing equipment are crucial because only the intentional alterations have to be compensated. The A/D transfer process should represent the original document characteristics, from either information and material points of view, as it arrived to us.

Fig. 11.1 summarizes the different points of view inside the debate evolved inside the archivist community on the audio documents re-recording.

According to the indications of the international archive community: 1) the re-recording is transferred from the original carrier; 2) if necessary, the carrier is cleaned and restored so as to repair any
Figure 11.1: The schema of the most significant positions of the debate evolved since the Seventies inside the archivist community on the audio documents active conservation.
climactic 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; 4) sampling frequency and bit rate must be chosen with respect to the archival sound record standard (see Sect. 11.4.3.1); 5) the digital audio file format should support high resolution, it should be transparent with simple coding schemes, without data reduction. Moreover, differently by Sch¨ uller position, it is our belief that - in a preservation copy - only the intentional alterations must be compensated (correct equalization of the re-recording system and decoding of any possible intentional signal processing interventions). All the unintentional alterations (also the ones caused by misalignments of the recording equipment) could be compensated only at the access copy level: these imperfections/distortions must be preserved because they witness the history of the audio document transmission.

Because these guidelines should be customized for each carrier, the archivists have to know all their implications, from physical and chemical points of view, and should possess a deep knowledge about the technology for re-recording and of the digital formats in which the digital preservation copy is to be stored.

### 11.3 Passive Preservation

The direct passive preservation can be carried only if the main causes of the physical Carriers deterioration are known and consequently avoided. We summarize the main risks for the two most common categories of carriers: mechanical carriers and magnetic tapes.

#### 11.3.1 Mechanical Carriers

The common factor with this group of documents is the method of recording the information, which is obtained by means of a groove cut into the surface by a stylus modulated by the sound, either directly in the case of acoustic recordings or by electronic amplifiers. Mechanical carriers include: phonograph cylinders; coarse groove gramophone, instantaneous and vinyl discs. Tab. 11.1 summarizes the typologies of these carriers.

The main causes of deterioration are related to the instability of mechanical carriers and can be summarized as:

1. **Humidity.** Humidity, as with all other data carriers, is a most dangerous factor. While shellac and vinyl discs are less prone to hydrolytic instability, most kinds of instantaneous discs are extremely...
endangered by hydrolysis. Additionally, all mechanical carriers may be affected by fungus growth which occurs at humidity levels above 65% RH.

2. **Temperature** Elevated temperatures beyond 40°C are dangerous, especially for vinyl discs and wax cylinders. Otherwise the temperature determines the speed of chemical reactions like hydrolysis and should therefore be kept reasonably low and, most importantly, stable to avoid unnecessary dimensional changes.

3. **Mechanical Deformation.** Mechanical integrity is of the greatest importance for this kind of carriers. It is imperative that scratches and other deformation caused by careless operation of replay equipment are avoided. The groove that carries the recorded information must be kept in an undistorted condition. While shellac discs are very fragile, instantaneous and vinyl discs are more likely to be bent by improper storage. Generally, all mechanical discs should be shelved vertically. The only exceptions are some soft variants of instantaneous discs.

4. **Dust and Dirt.** Dust and dirt of all kinds will deviate the pick-up stylus from its proper path causing audible cracks and clicks. Fingerprints are an ideal adhesive for foreign matter. A dust-free environment and cleanliness is, therefore, essential.

### 11.3.2 Magnetic Tape

The basic principles for recording signals on a magnetic medium were set out in a paper by Oberlin Smith in 1880. The idea was not taken any further until Valdemar Poulsen developed his wire recording system in 1898. Magnetic tape was developed in Germany in the mid 1930’s to record and store sounds. The use of tape for sound recording did not become widespread, however, until the 1950’s. Magnetic tape can be either reel to reel or in cassettes. Tab. 11.2 summarizes the typology of these supports:

<table>
<thead>
<tr>
<th>Period</th>
<th>Type of recording</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1935-1960</td>
<td>Analogue</td>
<td>base: cellulose acetate magnetic pigment: $Fe_2O_3$ formats: open reel</td>
</tr>
<tr>
<td>1944-1960</td>
<td>Analogue</td>
<td>base: PVC magnetic pigment: $Fe_2O_3$ formats: open reel</td>
</tr>
<tr>
<td>1959-</td>
<td>Analogue</td>
<td>base: polyester magnetic pigment: $Fe_2O_3$ formats: open reel</td>
</tr>
<tr>
<td>1969-</td>
<td>Analogue/Digital</td>
<td>base: polyester magnetic pigment: $CrO_2$ formats: compact cassette IEC II, DCC</td>
</tr>
</tbody>
</table>

**Table 11.2: Typology of magnetic tape carriers**

The main causes of deterioration are related to the instability of magnetic tape carriers and can be summarized as:
1. **Humidity.** Humidity is the most dangerous environmental factor. Water is the agent of the main chemical deterioration process of polymers: hydrolysis. Additionally, high humidity values (above 65% RH) encourage fungus growth, which literally eats up the pigment layer of magnetic tapes and floppy disks\(^4\) and also disturbs, if not prevents, proper reading of information.

2. **Temperature.** Temperature is responsible for dimensional changes of carriers, which is a particular problem for high density tape formats. Temperature also determines the speed of chemical processes: the higher the temperature, the faster a chemical reaction (e.g., hydrolysis) takes place.

3. **Mechanical Integrity.** Mechanical integrity is a much underrated factor in the accessibility of data recorded on magnetic media: even slight deformations may cause severe deficiencies in the playback process. Most careful handling has to be exercised, along with regular professional maintenance of replay equipment, which, in case of malfunctioning, can destroy delicate carriers such as R-DAT very quickly. With all tape formats, it is most important to obtain an absolutely flat surface of the tape pack to prevent damage to the tape edges which serve as mechanical references in the replay of many high density formats. All forms of tape should be stored upright.

4. **Dust and Dirt.** Dust and dirt prevents the intimate contact of replay heads to the medium which is essential for the correct access to the information especially with high density carriers. The higher the data density, the more cleanliness has to be observed. Even particles of cigarette smoke are big enough to hide information on modern magnetic formats. Also pollution caused by industrial smog can accelerate chemical deterioration. The effective prevention of dust is an indispensable measure for the proper preservation of magnetic media.

5. **Magnetic Stray Fields.** Magnetic stray fields are the natural enemy of magnetically recorded information. Sources of dangerous fields include dynamic microphones, loudspeakers and headsets. Also the simple magnets used for magnetic notice boards possess magnetic fields of dangerous magnitudes. By their nature, analogue audio recordings, including audio tracks on video tapes, are the most sensitive to magnetic stray fields. It should be noted that normally a distance of 10-15 cm is enough to diminish the field strength of even strong magnets to acceptably low values.

Among the others, some effects can be:

- “drop out” (i.e. the magnetic material fall off the tape);
- “bleed through” (i.e. the signal from one section of tape imprinting on another when the tape has been stored for a long time: this is a big issue in several magnetic recordings and is really noticeable in the excerpts with a low SNR);
- “stretch” (i.e. the actual permanent stretching of the polyester cause by too tightly spooling the tape with noticeable pitch dropping).

Tab. 11.3 shows the correct parameters for the passive preservation of mechanical and tape carriers.

### 11.4 Active Preservation

This section details a protocol for the task of audio documents active preservation, which is summarized in Fig. 11.2. The protocol has been defined by the author and put it into practice in several European audio archives projects (see Sect. 11.8).

\(^4\)Floppy disks are one of the most used supports to store audio documents in the field of electronic music in the 80s and 90s of the last century. The composers usually saved in floppy disks some short sound objects, synthesized at low sampling Hertz (8 – 15kHz). The study of this musical excerpt is very important from a musicologist point of view. For instance, the Archive of the Centro di Sonologia Computazionale (CSC, University of Padova, Italy: [http://csc.dei.unipd.it/](http://csc.dei.unipd.it/)) has hundreds of floppy disks: it is unquestionably an outstanding testimony of the musical history in the 80’ and 90’ years of XX Century.
11.4.1 Carrier Analysis and Restorative Actions

During this phase (steps 1 and 2 shown in Fig. 11.2) the state of the document must be evaluated and the physical characteristics of the carrier and its format assessed, also on the basis of historical research carried out on the technologies in use at the time of the recording. The preservative re-recording operation should be monitored so to memorize every phase of the process and to testify the accuracy of the protocol used. In particular, a video recording, synchronized with the audio signal, should document the presence of splices, corruptions and graphical signs. The documentation of this meaningful editing traces is very important for the signal alteration classification and for the philological work of genesis reconstruction.

The information on the format of the carrier has to be inferred from the direct analysis of the carrier and then compared with the technical data contained on the case/cover/label, even if it is often wrong or missing. The data inferred from the history of audio technology are a source of knowledge which cannot be ignored when defining methods and procedures for the survey of the formats and replay parameters adopted during the original recording, because they allow us to solve specific problems caused by the technical defects of the equipment used for the creation of the document. Clearly, all the results of this recognition have to be stored as additional information.
11.4.2 Re-recording

This phase details steps 3 and 4 shown in Fig. 11.2. On the basis of the information gathered in the first phase, the playback analogue equipment is chosen to avoid introducing further distortions and to collect more information than the one offered by the equipment of the time. The technical-functional analysis confirms the importance of this choice. For instance, tape recorders built before the 80s present:

a) low signal-to-noise-ratio (SNR);

b) fixed and non-modifiable equalizations;

c) unreliability of the tape transport system in guaranteeing the physical integrity of the original document. According to the considerations given in Sect. 11.2, the transfer from the old to the new format has to be carried out without subjective alterations or “improvements”, such as de-noising, because the unintended and undesirable artifacts are also part of the sound document, even if they have been subsequently added to the original signal by mishandling, poor storage or as a consequence of aging. Both have to be preserved with the utmost accuracy, because they provide information about the persons and the corporate bodies that were involved in the creation and in the transmission of the document. Alteration removal or attenuation on the signal need subjective choices of the restorer.

The A/D transfer is a delicate aspect of the re-recording procedure. Because original carriers may contain secondary information (i.e., bias frequency\(^5\), broadband impulsive noise) which falls outside the frequency range of the primary information (signal), the transfer must be carried out to the highest among the available standards.

Every audio document presents original technical aspects. It is precisely because of this instability inherent in the document that it is impossible to carry out automatic re-recordings with the simultaneous use of several systems. The process should be constantly monitored, and a number of signal alterations need to be catalogued and described:

- local noise: clicks, pops, signal dropout due to joints or tape degradation;
- global noise: hums, background noise, distortion (periodical or non-periodical);
- alterations produced during the sound recording phase: electrical noises (clicks, ripples), microphone distortions, blows on the microphone, induction noise;
- signal degradation due to malfunctions of the recording system (i.e., partial tracks deletion).

11.4.3 Preservation Copy

This section describes steps from 5 to 8 shown in Fig. 11.2.

A preservation copy (or archive copy) is the artifact designated to be stored and maintained as the preservation master. Such a designation may be given either to the earliest generation of the artifact held in the collection, to a preservation transfer copy of such an artifact, and/or to both such items in the possession of the archive. Such a designation means that the item is used only under exceptional circumstances\(^6\). During the process of active preservation, the original document – multimedia in itself, because is made up of the audio signal, static images (label, case, carrier corruptions, etc.), text (attach-
ments), smell (mould, vinegary, etc.) – is converted into a digital document, which could be defined as an unimedia document, because it is a fusion of different media in a single bit flow.

This projection of a multidimensional object into a one-dimensional space produces a particularly large and various set of digital documents, which are made up of the audio signal, the metadata and the contextual information. It is important to note that in this context, as it is common practice in the audio processing community, we use the term metadata to indicate content-dependent information that can be automatically extracted by the audio signal; as already mentioned we indicate as contextual information the additional content-independent information. The goal of active preservation is to minimize the information loss during the A/D transfer of the document. 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 the edition containers (envelopes, cases and boxes), on the label, on the flange, on the carrier and on possible attachments (text, images, physical conditions, intentional alterations, corruptions) and the information related the process of audio signal transfer (schemes of the A/D system) must be arranged and so they become a complete part of the conservative copy.

As for all types of digital documents, also in this case digital preservation methods and techniques have to be exploited, to maintain the accessibility of the preservation copy, its metadata and contextual information.

11.4.3.1 Format for the Audio Files

According to the rule the worse the signal, the higher the resolution, the audio signal should be stored in the preservation copy using the Broadcast Wave Format, sampled at least at 96 kHz with a 24 bit resolution. It is advisable to use the monophonic format, where each recording track is equivalent to a different file with Pulse Code Modulation representation.

In order to preserve sound documents in a philologically correct way during the re-recording procedures, it is essential to rely on operational protocols aimed at avoiding the overlapping of modern phonic aspects that alter the original sound content. In particular, the criteria for the preservation of documents should not be influenced by the market-induced tendency to use lossy compression formats. The low quality of lossy compression, especially if considered in relation to the phonic richness of much contemporary music, imposes the rigorous avoidance of any mixture between the acquisition of documents for conservative aims (preservation copies) and the archiving for common use (access copies).

11.4.3.2 Video Shooting and Photographic Documents

The information written on edition containers, labels and other attachments should be stored with the preservation copy as static images (two examples are given in Fig. 11.3 (a) and (b)), 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 recording offers:

1. Information related to magnetic tape assembly operations and corruptions of the carrier (disc, cylinder or tape), which are indispensable to distinguish the intentional from the unintentional alterations during the restoration process.

2. A description of the irregularities in the playback speed of analogue recordings (wow and flutter7):

---

7Wow and flutter are audio distortions perceived as an undesired frequency modulation in the range of: i) wow from 0.5 Hz to 6 Hz, ii) flutter from 6 Hz to 100 Hz. The distortions are introduced to a signal by an irregular velocity of the analogue medium. As the irregularities can originate from various mechanisms, the resulting parasitic frequency modulations can range from periodic to accidental, having different instantaneous values.
Figure 11.3: (a) a sound postcard: it looked like a standard postcard on the back, but on the front an analogue recording was engraved in a thin layer of laminate. Sound postcards were usually made by small firms, and the recording quality was extremely low; in this case the importance of storing the picture in with the preservation copy is particularly evident. (b) displays a label of His Master’s Voice disc: DK 119 (on the label, right) is the catalogue number; 2-054042 (on the label, left, and at the top of the mirror) is a second catalogue number (as its minor typographic importance, probably it is the first issue catalogue number: therefore here we have a reprint); A12804 (in the mirror, down) is the matrix number. It is possible to decode this information: DK = 30 cm diameter; Yellow label = “International Celebrità” series, printed in Hayes; 2-054 prefix in catalogue number corresponds to a second issue (2), 30 cm diameter (0), Italian catalogue (5) and duet or trio as sound content (4); by means of a comparison between matrix number and published repertories we can deduce the recording date (17th, January, 1913). (c) and (d) show two typical corruptions in a tape and in a disc respectively: this information should be stored with the preservation copy also, in order to have a deep insight the artifacts of the audio signal.
in discs, a spindle hole not precisely centered and/or the warping of the disc cause a pitch variation; in tape recorders, an irregular tape motion during playback (a change in the angular velocity of the capstan, or dragging of the tape within an audio cassette shell) cause changes in frequency. From the video, it is possible to locate automatically the imperfections occurred during the A/D transfer (see Sect. 11.5 for some examples): in this way, in the restoration process we will be able to distinguish among the alterations occurred at the recording step or at playback level.

3. Instructions for the performance of the piece (in particular in the electro-acoustic music for tape): from the video analysis, some prints of the tape can be displayed; they represent either the synchronization of the score or the indication of particular sound events (Fig. 11.4).

**Figure 11.4:** Frame of a video recording of an open reel tape: the circle drawn in black marks a specific sound event. Often, in the electro-acoustic music field (in the works for tape and acoustic musical instruments) the marks on the tape are used as a synchronization means between live-electronics performer and the recorded tape music. If this information was not preserved, it would not be possible to perform the piece.

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 our experience, a 320x240 pixels resolution video with medium quality DivX compression yielded satisfactory results.

### 11.4.3.3 Audio Fingerprinting

The deterioration of the digital carrier used for storing the preservation copy could cause some errors in the audio files. If the errors are restricted to the bits assigned to the audio signal codification, however the file is proved to be readable, but it is no longer capable of returning exactly an audio signal equal to the
one which was digitized. A control device of the integrity of the audio files, thus, should be introduced in the preservation copy.

A common approach to face this problem is the use of error detection codes, for instance hashing techniques such as MD5 that are computed over the complete file and help identifying changes in the bit flow. In order to highlight the actual temporal positioning of these changes, we propose to enrich the metadata extracted from images and videos of the carrier with an audio fingerprint of the audio signal. A fingerprint is a unique set of features automatically extracted from the audio signal that aims at the identification of digital copies, even in presence of noise, distortion, and compression. To this end, a fingerprint can be considered as a content-based signature that summarizes an audio recording. It is important to note that, although robust to noise, typical audio fingerprinting techniques can measure the difference between the original signal and the distorted copies.

Although usually aimed at digital rights management, being a compact representation of the audio signal, fingerprinting can find useful applications also in the development of music digital libraries other than tracking the diffusion of illegal copies of protected material. In particular, it can be useful to align different audio files of the same re-recording procedure, for instance the high quality audio which is the main goal of the A/D conversion and the low quality audio embedded in the video capture. Moreover, periodic extraction and comparison of the fingerprints can detect the exact time positioning of errors in the preservation copy due to aging of the digital carrier. Finally, we propose that fingerprinting can be used to measure the difference between the preservation and the access copies, because they are both originated from the same audio file.

Another technique that is worth mentioning, and which is often considered an alternative to audio fingerprinting, is audio watermarking. In this case, research on psychoacoustics is exploited to embed in a digital recording an arbitrary message, the watermark, without altering the human perception of the sound. The message can provide contextual information about the recording (such as title, author, performers), the copyright owner, and the user that purchases the digital item. Also in this case, this latter information can be useful to track the responsible of an illegal distribution of digital material. Similarly to fingerprints, audio watermarks should be robust to distortions, additional noise, A/D and D/A conversions, and compressions. Yet, the message that can be inserted through non-audible watermarking is still limited, and thus this technique cannot be used for embed complex information into the signal. Surely, audio watermarking should be used to add a unique identifier at least to any access copy.

11.4.3.4 Descriptive card

The data stored in the preservation copy can be easily copied onto new digital carriers. As digital carriers have a incremented data storage capacity (optical carriers: HD-DVD and Blue-ray Disc up to 50 GB of storage; cartridge digital magnetic tapes up to 800 GB of storage; HDD with some TB), the modification of the data organization in the preservation copies is to be expected, by introducing more documents into the same carrier and by adopting a different logical structure. For this purpose, it is necessary to provide the preservation copy with a list of all the documents belonging to the preservation copy, some metadata of the audio signal, and a description of the analogue original document. In our experience, it is necessary a descriptive card composed, at least, by five elements:

1. Heading;
2. Description of the preservation copy;
3. List of the documents stored in the preservation copy;
4. Description of the original document;
5. Description of the video recording.
11.5 Automatic Metadata Extraction

The increased dimensionality of the data contained within an audio digital library, which has been explained in the previous section, should be dealt with by means of automatic annotation. The auditory information contained in the audio medium can be augmented with cross-modal cues. For instance, the visual and textual information carried by the cover, the label and other attachments should be acquired through photos and/or videos. The extraction of this valuable information can be performed through well-known techniques for image and video processing, such as OCR, video segmentation and so on. We believe that it is interesting as well, even if not studied yet, to deal with other visual information regarding the carrier corruption and imperfection occurred during the A/D transfer.

Computer vision algorithms and techniques can be applied to the automatic extraction of relevant metadata. This section presents a set of tools able to extract, automatically, metadata from photos and video recordings of magnetic tape and phonographic disc.

11.5.1 Reel to Reel Magnetic Tape

The auditory information contained in the audio medium can be augmented with cross-modal cues. For example, a video of a winding tape can document its state of preservation and record precious information such as the presence of splices and marks. Regarding video, well-known techniques such as change detection by background subtraction can be applied to detect discontinuities as seen in Fig. 11.5. In this case, I have employed background subtraction with automatic thresholding and a voting step to detect major changes in the image due to the presence of different materials (i.e., magnetic vs. header tapes).

![Figure 11.5](image)

**Figure 11.5:** (a) and (b) show source frames from the video of a winding tape, while (c) and (d) show the corresponding processed images.

Fig. 11.5(c) is completely black as no significant changes have been detected between the current frame of Fig. 11.5(a) and the background image. In Fig. 11.5(d) a major change has occurred (white pixels) in the source frame shown in Fig. 11.5(b) (tape without magnetic layer). Therefore, the automatic detection of the start of a magnetic tape can be performed in a very simple and effective way via the processing steps mentioned above and by setting a threshold on the percentage of changed pixels with respect to the Region Of Interest (ROI). The ROI could be set in order to focus the algorithms only on a subregion of the image. As it can be seen in the source frames of Fig. 11.5, the tape occupies roughly 50% of the image, while other details such as the player’s heads are not relevant for the processing and should be discarded by setting a ROI on the tape region. The approach described above is very similar to the techniques used for scene cut detection for automatic annotation of video sequences.

Fig. 11.6 shows how other information can be extracted by processing the videos of a winding tape. The basic processing steps are the same employed in the previous experiment, additional steps are required to detect splices or specific marks. In Fig. 11.6(b) no significant changes are detected, the image
is not completely black but detected changes do not form a connected component large enough to pass
the threshold.

Fig. 11.6(d) shows how a tape splice can be detected. The Hough transform is applied to detect
lines in the subregions where changes have been detected. As it can be seen, the transform detects a
line corresponding to the tape splice. In Fig. 11.6(f) a connected component corresponding to the dot
in Fig. 11.6(e) is detected. The system can therefore annotate the corresponding frame linking it to the
specific sound event marked by the felt-tip pen sign.

11.5.2 Warped Phonographic Discs

The characteristics of the arm’s oscillations can be related to pitch variation of the audio signal. As such,
they constitute valuable metadata for audio signal restoration processes. Also in this case, computer
vision techniques can be applied to the automatic analysis of rotating discs. We have employed a feature
tracking algorithm known as the Lucas-Kanade tracker. The algorithm locates feature points on the image
to be tracked between consecutive frames. The technique, initially conceived for image registration, is
here employed as a feature tracker to keep track of the position of the features from a frame to the
following one. Fig. 11.7 shows some frames from one of the sequences used in the experiments: (b)
shows the lowest position of the arm’s head in one oscillation and (c) the highest position, where the
Lucas-Kanade features can be seen on the arm’s head while being tracked through the oscillation. Even
if from the Fig. 11.7 the differences between the highest and lowest positions are almost unnoticeable (see
the differences between them in (d)), our approach is able to track them clearly, as shown in Fig. 11.8.

Fig. 11.8 shows the temporal evolution of the y coordinate of a feature located on the arm’s head. The
x-axis shows the number of frames and the y-axis reports the position in pixels on the image plane. The
oscillatory evolution is clearly visible. There is a 29 frames gap between Fig. 11.7(b) and Fig. 11.7(c),
which is consistent to the period of the oscillations shown in Fig. 11.8.

11.5.3 Off-centered Phonographic Disc

Interesting properties of a phonograph record can be automatically extracted by analyzing a picture of it.
For example, we wanted to calculate the eccentricity of the disc, that is, the offset between the spindle
hole axis and the exact central rotation axis. This production flaw, which could affect individual copies
or entire stocks of records, is responsible for the well-know warp effect that introduces a pitch variation
in the audio signal. To accomplish this automatically I have exploited the consolidated literature on iris
detection. Since our problem shares the same lucky circular properties of the problem of iris detection, we
have employed the integrodifferential operator which was developed for detecting the pupillary boundary
and the outer boundary of the iris. The integrodifferential operator has the following form:

$$
\max_{(r,x_0,y_0)} \left| \frac{\partial}{\partial r} \int_{r,x_0,y_0} \frac{I(x,y)}{2\pi r} ds \right| 
$$

The operator is computed over the image \(I(x,y)\) where it searches for the maximum of the blurred
partial derivative, with respect to the radius \(r\), of the normalized circular integral of radius \(r\) and center
coordinates \((x_0, y_0)\) calculated on \(I(x, y)\). The blur is obtained through convolution with a Gaussian
smoothing function of scale \(\sigma\). In other words, the operator works as circular edge detector and provides
the centre coordinates and the radius of the strongest circular edge in the image. The outer contour of the
disc is extracted and then the operator on the image for detecting the spindle hole contour is rerun The
second pass can be computed very fast as it takes advantage of the known geometrical properties of vinyl
discs. That is, once the outer boundary has been detected the spindle hole contour can be searched in a
subregion of the image inside the outer contour. The disc was laying on a plane parallel to the image and
Figure 11.6: Automatic discontinuities extraction from a winding tape (splices, marks).
Chapter 11. Musical cultural heritage: From preservation to restoration

Figure 11.7: Processed frames from a video of a oscillating record player’s arm. (a) Photo of the turntable arm; (b) Lowest position of the arm in an oscillation, (c) its highest position. (b) and (c) show Lucas-Kanade features detected on the arm’s head and tracked through the oscillation. (d) shows the differences between lowest and highest positions.

Figure 11.8: Temporal evolution of the y coordinate of a Lucas-Kanade feature located on the arm’s head. It can be seen clearly how the oscillations indicate a deformed disc.
the spindle hole was on-axis with the camera’s optical axis. Although this constraint is not particularly restrictive for a dedicated set-up in an audio laboratory, a step further can be taken by removing this assumption and considering perspective deformations given by out-of-axis images.

Having detected the outer boundary of the disc and the spindle hole contour, the calculation of the offset between their centers is trivial. In the author experience, the estimated offset can be greater then 1 cm. The processing described in this subsection can be performed on-line in real-time. The experiments shown in Fig. 11.5, Fig. 11.6 and Fig. 11.7, have been carried out on off-line 320x240 resolution video sequences with an above real-time frame rate processing performance of 50 frames/sec on a 3 GHz single processor machine. The application has been coded in C++. In addition, no particular setup was required for this experiment. Video sequences have been acquired with a consumer digital camcorder at PAL resolution and subsequently rescaled and compressed into DivX video files at medium-high quality setting. As can be seen comparing Fig. 11.5, Fig. 11.6 and Fig. 11.7, the algorithms are robust to different lighting conditions. The achieved results hint the possibility to perform tape marks detection in real-time, as the tape is winding. This would be a practical set-up for audio laboratories and audio digital libraries.

11.5.4 Representing Metadata

Once all this content-dependent information has been extracted, a suitable metadata schema for its representation has to be chosen for its representation. Among the existing metadata standards, probably the Metadata Encoding and Transmission Standard (METS) is particularly suitable for representing the information about the carriers and the A/D transfer. It can be noted that METS has already been used to encode music documents with profiles for both scores and sound recordings, for instance in the Digital Library of the Brown University. The, METS documents have two sections that are particularly significant for the aims of this study: the File Section allows us to keep information about additional files, which is particularly significant since also the extracted metadata is in the form of additional multimedia documents, and the Structural Map that can represent the hierarchy between different metadata, for instance ranging from the the video capture of the A/D transfer of a warped phonographic disc, to the tracking of feature points on the pickup, to the representation of the movement of the pickup along the vertical axis, as explained in Sect. 11.5.2.

As it is well-known, another suitable schema for music documents is MPEG. In particular, MPEG-7 can easily represent the description, the definition and the content of extracted metadata as accompanying features of the audio digital object. The application of MPEG-7 seems particularly appealing because of its ability to describe low-level characteristics, as the ones extracted automatically from the images of the carrier and the video of the A/D transfer. The XML-based structure of MPEG-7 allows a straightforward extension to include the multimedia material and the results of the analysis techniques presented in this and in the following sections. Yet, a discussion of the metadata schema is beyond of the scope of this paper.

11.6 Audio Data Extraction and Alignment from Phonographic Disc

This section introduces: a) a system for reconstructing the audio signal from a still image of a phonographic disc surface; b) alignment techniques useful in the comparison of alternative digital acquisitions. A case study where the alignment tool is used to annotate disc corruptions is described in the following section.
11.6.1 Photos of GHOSTS (PoG)

Nowadays, automatic text scanning and optical character recognition are in wide use at major libraries. Yet, unlike text scanning, A/D transfer of historical sound recordings is often an invasive process.

As it is well-known, several phonographs exist that are able to play gramophone records using a laser beam as pickup (laser turntable). This playback system has the advantage of never physically touch the record during playback: the laser beam traces the signal undulations in the record, without friction. Unfortunately, the laser turntables are constrained to the reflected laser spot only and are susceptible to damage and debris and very sensitive to surface reflectivity.

Digital image processing techniques can be applied to the problem of extracting audio data from recorded grooves, acquired using a digital camera or other imaging system. The images can then be processed to extract audio data. Such an approach offers a way to provide non-contact reconstruction and may in principle sample any region of the groove, also in the case of a broken disc. These scanning methods have several advantages: a) delicate samples can be played without further damage; b) broken samples can be re-assembled virtually; c) the re-recording approach is independent from record material and format (wax, metal, shellac, acetates, etc.); d) effects of damage and debris (noise sources) can be reduced through image processing; e) scratched regions can be interpolated; f) discrete noise sources are resolved in the “spatial domain” where they originate rather than being an effect in the audio playback; g) dynamic effects of damage (skips, ringing) are absent; h) classic distortions (wow, flutter, tracking errors, etc) are absent or removed as geometrical corrections; i) no mechanical method is needed to follow the groove; l) they can be used for mass digitization.

In the literature, there are several approaches to this problem, based on: Digital Cameras (2D or horizontal only view, frame based); Confocal Scanning (3D or vertical+horizontal view, point based); Chromatic sensors (3D, point based); White Light Interferometry (3D, frame based). The authors have developed the Photos of GHOSTS (PoG) system that: a) is able to recognize different rpm and to perform track separation automatically; b) does not require human intervention; c) works with low-cost hardware; d) is robust with respect to dust and scratches; e) outputs de-noised and de-wowed audio, by means of novel restoration algorithms. The user can choose to apply an equalization curve among the hundreds stored in the system, each one with appropriated references (date, company, roll-off, turnover). Moreover, PoG allows the user to process the signal by means of several audio restoration algorithms. The software automatically finds the record centre and radius from the scanned data, for groove rectification and for track separation. Starting from the light intensity curve of the pixels in the scanned image, the groove is modeled and the audio samples are obtained. The complete process is depicted in Fig. 11.9.

The system enhancements include:

1. the user can select the correct equalization in a list including 225 different curves, able to cover all the electric recordings, since 1925.

2. A de-noise algorithm in a frequency domain based on the use of a suppression rule, which considers the psychoacoustics masking effect. The spreading thresholds which present the original signal $x(n)$ are not known a priori and are to be calculated. This estimation can be obtained by applying a noise reduction STSA standard technique leading to an estimate in the frequency domain.

Audio restoration algorithms can be divided in three categories:

(a) frequency-domain methods, such as various forms of noncasual 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 (only the Power Spectral Density noise estimation);

(b) time-domain restoration by signal models such as Extended Kalman filtering: in these methods it is necessary a lot of a priori information in order to estimate the statistical description of the audio events;

(c) restoration by source models: only a priori information is used.
of $x(n)$, for which the masking thresholds $m_k$, defined as the non negative threshold under which the listener does not perceive an additional noise, can be calculated by using an appropriate psychoacoustic model. The masking effect obtained is incorporated into one of the EMSR technique, taking into consideration the masking thresholds $m_k$ for each $k$ frequency of the STFT transform. A cost function depending on $m_k$, which minimization gives the suppression rule for the noise reduction, is created. This cost function can be a particularization of the mean square deviation to include the masking thresholds, under which the cost of an error is equal to zero.

3. The design and the realization of ad-hoc prototype of a customized scanner device with a rotating lamp carriage in order to position every sector with the optimal alignment relative to the lamp (coaxially incident light). In this way we improved (from experimental results: more than 20%) the accuracy of the groove tracking step.

PoG may form the basis of a strategy for: a) larger scale A/D transfer of mechanical recordings which retains maximal information (2D or 3D model of the grooves) about the native carrier; b) small scale A/D transfer processes, where there are not sufficient resources (trained personnel and/or high-end equipments) for a traditional transfer by means of turntables and converters; c) the active preservation of carriers with heavy degradation (breakage, flaking, exudation).

11.7 Audio restoration

The audio restoration algorithms can be divided into three categories:

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: 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 $\lambda$, the forgetting factor $\gamma$, the threshold $\mu$ for detection of impulsive noise).

This section presents algorithms, developed at the Centro di Sonologia Computazione (Dept. Information Engineering) using the VST plug-in architecture, able to offer satisfying examples of the above mentioned categories. The algorithms are detailed in the next subsections:

- **CREAK** (Canazza REstoration Audio - extended Kalman filter): A de-noise and de-click system based on Extended Kalman Filter, dedicated to the restoration of audio signal re-recorded from shellac discs: low Signal to Noise Ratio (SNR), clicks, pops, crackle.
- **CMSR** (Canazza-Mian Suppression Rule): 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.
- **PAR** (Perceptual Audio Restoration): A de-hiss based on perceptual algorithm for reel-to-reel tapes and cassettes: high SNR.
Figure 11.9: Photos of GHOSTS schema.
Of course, regardless their dedications, in a real restoration work it is opportune to combine these tools in order to obtain the better results.

11.7.1 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 Extended Kalman Filter theory (EKF), as proposed by M. Niedzwiecki and K. Cisowski. In particular the algorithm 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.

11.7.1.1 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’ve 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.

11.7.1.2 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.

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.

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:

\[
\hat{s}_w(t) = \frac{\hat{s}_+^2(t)}{\hat{s}_+^2(t) + \hat{s}_-^2(t)} \hat{s}_+(t) + \frac{\hat{s}_-^2(t)}{\hat{s}_+^2(t) + \hat{s}_-^2(t)} \hat{s}_-(t)
\]

(11.2)
Chapter 11. Musical cultural heritage: From preservation to restoration

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

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.

11.7.2 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: 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. A substantial progress was made with the solution carried out in Ephraim and Malah, 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 \( Y_{\text{prio}} \) (SNR calculated taking into account the information of the preceding frame) and \( Y_{\text{post}} \) (SNR calculated taking into account the information of the current frame). A parameter \( (\alpha) \) 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. \( Y_{\text{prio}} \) has less variance than \( Y_{\text{post}} \): this way, musical noise is less likely to occur.

Unfortunately, in the case of cylinders or shellac discs an optimal value of \( \alpha \) 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\(^9\) Suppression Rule, CMSR), based on the idea of using a punctual suppression without memory (Wiener like) in the case of a null estimate of \( Y_{\text{post}} \), according to:

\[
\alpha = \begin{cases} 
0.98, & \text{if } Y_{\text{post}}(k,p) > 0 \\
0, & \text{otherwise}.
\end{cases} \quad (11.3)
\]

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

\(^9\)Gian Antonio 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.

This book is licensed under the CreativeCommons Attribution-NonCommercial-ShareAlike 3.0 license, ©2005-2013 by the authors except for paragraphs labeled as adapted from <reference>
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.

11.7.3 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).

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, sketched in Figure 11.10. 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 $Y_{\text{prio}}$ and $Y_{\text{post}}$ terms (see Sec. 11.7.2) are calculated according to the inner representation and the gain $G(p, z)$ is derived. Figure 11.11 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.

![Figure 11.10](image)

Figure 11.10: 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.

11.7.4 Experimental results

A series of experiments with real usage data from different international audio archives were conducted. In this section experimental results of applying the above described techniques related to audio restoration are presented. As first case study, Figure 11.12 shows a restoration of a wax cylinder by means of CMSR (see Sec. 11.7.2). 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 Figure 11.12: 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 Figure 11.13 shows a de-click of a shellac disc by means of CREAK (see Sec. 11.7.1). The song is La signorina sfinciusa (The funny girl), performed by Leonardo
Figure 11.11: 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).

Figure 11.12: 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).
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, 24th, 1929. In Figure 11.13 is pointed out a click, before (top) and after (bottom) the audio restoration performed by CREAK.

![Figure 11.13: 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).](image)

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) is considered. In this case we performed only a de-noise by means of PAR tool. Figures 11.14 and 11.15 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, 11th, 1919. In this case, we carried out de-click and de-noise by means of CREAK. Because of the low SNR (SNR ∼ 5 dB), we processed the signal also with CMSR. In this way, we obtained a SNR = 40 dB, without introducing particularly audible distortions (musical noise). Figures 11.16 and 11.17 show the corrupted (top) and restored (middle and bottom) signals respectively in the time and frequency domains.

### 11.7.4.1 Comparison

Figure 11.18 shows the gain trend introduced by the filters described above in comparison with some standard filters (Wiener filter, Power Subtraction, EMSR) at the varying of the noisy signal SNR, con-

---

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

11 The measuring of noise power is made by taking the noise print in an interval where there is only background noise.
Figure 11.14: Top: the waveform of the original (corrupted) audio extract. Bottom: the restored data by PAR. X-axis: time (s). Y-axis: amplitude (normalized).

Figure 11.15: Top: the spectrum of the original (corrupted) audio extract. Bottom: the restored data by PAR. X-axis: time (s). Y-axis: frequency (Hertz).
Figure 11.16: 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).

Figure 11.17: 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 (Hertz).
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 $[15 \div 30]$ dB and PAR seems adapt to reduce the noise in audio signal with a good SNR.

![Figure 11.18: 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 $[15 \div 30]$ dB; PAR for high SNR.]

### 11.7.5 Assessment

In order to show the methodology usually adopted in the field of audio enhancement and compression (e.g., by the MPEG group) to validate audio processing systems, a listening test carried out to validate the audio restoration algorithms detailed above is described.

**Material.** Three music pieces recorded in clearly different supports were used (see Sec. 11.7.4 for the details of these audio documents):

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 seconds 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 able 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.

Test method. The tests were conducted using the EBU MUSHRA test method, which is a recommended evaluation method adopted by ITU. 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 kHz and 3.5 kHz).

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.

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.

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.

Test duration. The training session for each listener took approximately 1 hour, 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 minutes. Subjects were allowed a rest period between each session, but not during a session.

Main results. The statistical analysis method described in the MUSHRA specification was used to process the test data. The results are presented in Tab. 11.4 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.

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.

Discussion. It is possible to make some important comments:
Chapter 11. Musical cultural heritage: From preservation to restoration

Table 11.4: Mean for restored stimuli and anchors, 24 subjects. Stimuli: S1 = My Mariuccia take-a steamboat; S2 = La signorina sfinciusa; S3 = Unpublished tape recording

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

- 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 11.7.4.1). 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. 11.4). However, let us consider the three tools CREAK, CMSR and PAR as a single restoration environment (it could be called CARE tool: CAzzazza REstoration or Csc Audio REstoration\textsuperscript{12}): 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.

11.8 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.

Musicologist, 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,

\textsuperscript{12}CSC stands for Centro Sonologia Computazionale, the international laboratory of the University of Padua, leader in the Sound and Music Computing field since 1979.
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.

The goal of this chapter is to stress that the archiving process of digitized audio documents is complete only when it includes all the ancillary information, in particular metadata of the original carriers. In this sense guidelines to the A/D transfer are detailed, in order to minimize the information loss and to automatically measure the unintentional alterations introduced by the A/D equipment. In addition, this chapter has presented:

1. A novel system able to synthesize the audio signal from a still image of a phonographic disc surface.

2. A software to extract metadata from photos and video shootings of audio carriers.

In relation to audio restoration, this chapter has presented:

- The CARE tool (Sec. 11.7). Sometimes, music audio documents are usually recorded in non-professional carriers by means of amateur recording system (e.g., ethnic music field). 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 (Sec. 11.7.4), 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 (Sec. 11.7.4.1).

- A perceptual assessment, using the EBU MUSHRA test protocol (Sec. 11.7.5).

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 (Sec. 11.7.5) 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 chapter summarizes a number of experiences in several research/applied project on Digital Audio Archives and Audio Access, 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 Access of Contemporary Music Italian Archive” funded by the Italian Ministry for Scientific Research; “Preservation and Online fruition of the Audio Documents from the European Archives of Ethnic Music” funded by the EU under the Program Culture2000; “Search in Audio-Visual Content Using Peer-to-Peer Information Retrieval” funded by the EU under the Sixth Framework Programme; “Restoration of the Vicentini Archive in Verona and its Accessibility as an Audio e-Library”, joint project between the University of Verona and Arena Foundation. Equally important for defining the protocols described in this paper, has been the collaboration with important European audio archives, including: “Speech and Music Archives” of the National Research Council of Italy “Archive of the Studio di Fonologia Musicale”, owned by the Italian National Broadcaster Television; “Luigi Nono Archive”; “Bruno Maderna Archive”; “Historic Archive of Contemporary Arts” of the LaBiennale of Venice.

11.9 Commented bibliography

References


