Int J Digit Libr (2009) 10:201–220
DOI 10.1007/s00799-010-0060-6

Methodologies and tools for audio digital archives

Nicola Orio · Lauro Snidaro · Sergio Canazza · Gian Luca Foresti

Published online: 4 July 2010

Abstract In response to the proposal of digitizing the entire back-run of several European audio archives, many research projects have been carried out in order to discover the technical issues involved in making prestigious audio documents digitally available, which are related to the A/D transfer process and supervised metadata extraction. This article gives an innovative approach to metadata extraction from such a complex source material. This article also describes the protocols defined, the processes undertaken, the results ascertained from several audio documents preservation projects and the techniques used. In addition, a number of recommendations are given for the re-recording process, aimed at minimizing the information loss and to automatically measure the unintentional alterations introduced by the A/D equipment.

Keywords A/D transfer · Metadata · Digital archives · Historical audio documents

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, which 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 cataloging 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 addition, the audio processing carried out by the Tonmeister[1] is a real interpretative element added to the

N. Orio · S. Canazza (✉)
Department of Information Engineering, University of Padova, Padova, Italy
e-mail: canazza@dei.unipd.it

N. Orio
e-mail: orio@dei.unipd.it

L. Snidaro · G. L. Foresti
Department of Mathematics and Informatics, University of Udine, Udine, Italy
e-mail: lauro.snidaro@dimi.uniud.it

G. L. Foresti
e-mail: gianluca.foresti@dimi.uniud.it

[1] The term Tonmeister describes a person who has a detailed theoretical and practical knowledge of all aspects of sound recording. However, unlike a sound engineer, he/she must be also deeply musically trained. Both competencies have equal importance in a Tonmeister's study [2].

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[2] on article 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 analog and digital approaches and carriers—paper, wire, wax cylinder, shellac disk, film, magnetic tape, vinyl record, magnetic and optical disk, 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[3] preservation, which is the defence of the carrier from external agents without altering the structure, and an active preservation, which involves data transfer on new media.

It is worth noting that, in the 1970s/1980S of twentieth century, expert associations (Audio Engineering Society: AES; National Archives and Records Administration: NARA; and 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 analog 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; and (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. Smith, still in 1999, suggested that digitization should be considered a means for access, not preservation—"at least not yet" [61].

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, analog holdings must be digitized. At the end of the twentieth century, the traditional "preserve the original" paradigm shifted to the "distribution is preservation" [24] 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 [3,4,9,14,59].

This article proposes an innovative approach to metadata extraction from audio documents. After a detailed overview of the debate evolved since the 1970S inside the archivist community on audio documents preservation (Sect. 2), the article describes the protocols defined, the processes undertaken, the results ascertained from several international audio documents preservation projects and the techniques used. In particular, in Sects. 3 and 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. We believe that the increased dimensionality of the data contained within an audio digital library should be dealt with by means of automatic annotations. Therefore, this study presents in Sect. 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. 6 presents an original system for reconstructing the audio signal from a still image of a disk surface and an alignment technique aimed at comparing the effectiveness and the robustness of different re-recording techniques. Finally, Sect. 7 provides two case studies in which an alignment tool is used to annotate disk corruptions.

2 Audio documents preservation

A reconnaissance on the most significant positions of the debate evolved since the 1970S inside the archivist community on the audio documents active conservation highlights at least three different points of view [53], described below.

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 focused on the problem of standardizing the procedures of audio restoration in an article which became famous for the numerous controversies it arose [63]. Storm individuated two legitimate directions, two

[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.

[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.

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." [63]. 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" [63], with the intent of obtaining "the live sound of original performers," transcending the limits of a historically faithful reproduction of the recording.

2.2 "To save history, not rewrite it"

The Guide [14] 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 [59] 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" [14]. 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 toward 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 [14].

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 errors caused by misaligned recording equipment and replayed on modern equipment to minimize replay distortions" [14].

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, Schüller 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" [59]. While the compensations of type B are commonly accepted and must—as Schüller 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" [59]. 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.

2.3 "Secondary information": the history of the audio document transmission

The studies of George Brock-Nannestad [15] 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 [48], 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 historic-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.

2.4 A proposal for an audio preservation protocol

Starting from these positions, we define a preservation copy a digital data set that groups the information carried by the audio document, considered as an artifact (see Sect. 3.4 for details). 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.

Figure 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 [3–5,7–9,37–39]: (1) the re-recording is transferred from the original carrier; (2) if necessary, the carrier is cleaned and restored so as to repair any 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. 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 Schüller position [59], 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 physic and chemical points of view, and should posses a deep knowledge about the technology for re-recording and of the digital formats in which the digital preservation copy is to be stored.

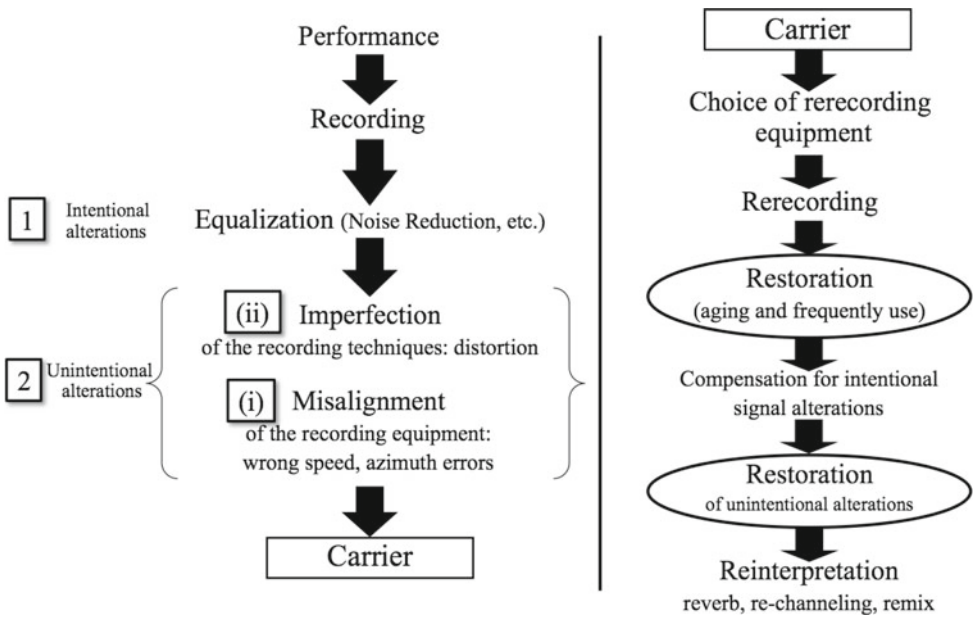

Fig. 1 The schema of the most significant positions of the debate evolved since the 1970S inside the archivist community on the audio documents active conservation

Table 1 Typologies of analogue mechanical carriers

Carrier	Period	Composition	*Stocks*
Cylinder—recordable	1886–1950s	Wax	300,000
Cylinder—replicated	1902–1929	Wax and Nitrocellulose with plaster ("Blue Amberol")	1,500,000
Coarse groove disk—replicated	1887–1960	Mineral powders bound by organic binder ("shellac")	10,000,000
Coarse and microgroove discs—recordable ("instantaneous discs")	1930–1950s	Acetate or nitrate cellulose coating on aluminum (or glass, steel, card)	3,000,000
Microgroove disk ("vinyl")—replicated	1948–	Polyvinyl chloride—polyacetate co-polymer	30,000,000

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.

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 disks. Table 1 summarizes the typologies of these carriers [17,18,39,40,42,57].

The main causes of deterioration are related to the instability of mechanical carriers and can be summarized as [18,39,42,57]:

1. *Humidity*. Humidity, as with all other data carriers, is a most dangerous factor. While shellac and vinyl disks are less prone to hydrolytic instability, most kinds of instantaneous disks are extremely endangered by hydrolysis. In addition, all mechanical carriers may be affected by fungus growth which occurs at humidity levels above 65% RH.
2. *Temperature*. Elevated temperatures beyond 40C are angerous, especially for vinyl disks and wax cylinders. Otherwise he temperature determines the speed of chemical reactions like ydrolysis and should, therefore, be kept reasonably low and, most mportantly, 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 disks are very fragile, instantaneous and vinyl disks are more likely to be bent by improper storage. Generally, all mechanical disks should be shelved vertically. The only exceptions are some soft variants of instantaneous disks.
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.

3.2 Magnetic tape

The basic principles for recording signals on a magnetic medium were set out in an article 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-1930s to record and store sounds. The use of tape for sound recording did not become widespread, however, until the 1950s. Magnetic tape can be either reel to reel or in cassettes. Table 2 summarizes the typology of these supports:

The main causes of deterioration are related to the instability of magnetic tape carriers and can be summarized as [11,18,32,39,41,57]:

1. *Humidity*. Humidity is the most dangerous environmental factor. Water is the agent of the main chemical deterioration process of polymers: hydrolysis. In addition, high humidity values (above 65% RH) encourage

Table 2 Typology of magnetic tape carriers

Period	Type of recording	Composition
1935–1960	Analogue	Base: cellulose acetate magnetic pigment: Fe_2O_3 formats: open reel
1944–1960	Analogue	Base: PVC magnetic pigment: Fe_2O_3 formats: open reel
1959–	Analogue	Base: polyester magnetic pigment: $_2O_3$ formats: open reel
1969–	Analogue/digital	Base: polyester magnetic pigment: CrO_2 formats: compact cassette IEC II, DCC
1979	Analogue/digital	Base: polyester magnetic pigment: metal particle formats: compact cassette IEC IV, R-DAT

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, analog 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).

Table 3 shows the correct parameters for the passive preservation of mechanical and tape carriers [18,39,42,57].

4 Active preservation

This section details a protocol for the task of audio documents active preservation, which is summarized in Fig. 2. The protocol has been defined by the authors and put it into practice in several European audio archives projects (see Sect. 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 (8–15 kHz). 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/) has hundreds of floppy disks: it is unquestionably an outstanding testimony of the musical history in the 80' and 90' years of twentieth century.

Table 3 Recommended climatic storage parameters for mechanical and tape characters

	Temp.	±/24 h	±/Year	RH	±/24 h	±/Year
Preservation storage	$5°C < t < 10°C$	±1°C	±2°C	30%	±5%	±5%
Access storage	About 20°C	±1°C	±2°C	40%	±5%	±5

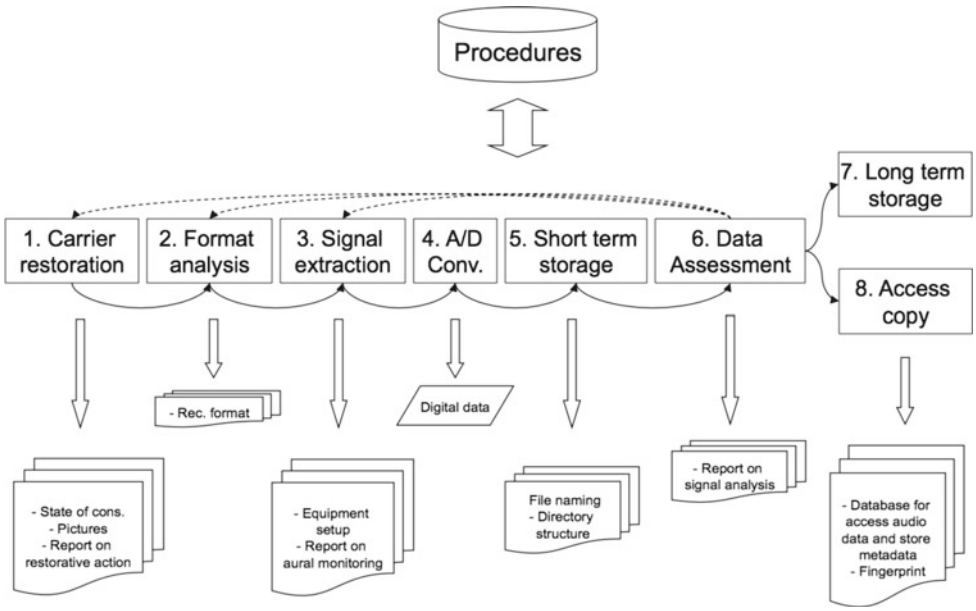

Fig. 2 Representation of the A/D transfer protocol

4.1 Carrier analysis and restorative actions

During this phase (steps 1 and 2 shown in Fig. 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.

4.2 Re-recording

This phase details steps 3 and 4 shown in Fig. 2. On the basis of the information gathered in the first phase, the playback analog 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. 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 cataloged 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).

4.3 Preservation copy

This section describes steps from 5 to 8 shown in Fig. 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]" [38]. 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 (attachments), smell (mold, 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 [51].

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.

4.3.1 Format for the audio files

According to the well-known 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 [36,37]. For further details on Broadcast Wave Format refer to [6,28].

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

[5] bias is the addition of an inaudible high-frequency signal to the audio signal. Bias increases the signal quality of audio recordings pushing the signal into the linear zone of the tape's transfer function [1].

[6] Audio carriers, especially modern high density formats, are, by their very nature, vulnerable. In addition, there is always the risk of accidental damage through improper handling, malfunctioning equipment or disaster. One strategy, for the long term storage, that is widely used is the creation of access copies of documents. A poor quality copy can act as an adjunct to the catalog to aid researchers to decide what documents they wish to study. A good quality copy may be acceptable for study in place of the original. The (online or local) use of copies to reduce the frequency of access to the original document will reduce the stress on the original and help to preserve it. A clear policy about the classes of researchers allowed access to original documents—particularly fragile ones—will also help documents survive. It is clearly impossible to totally restrict access to originals but many users can perform their research using good quality access copies [39].

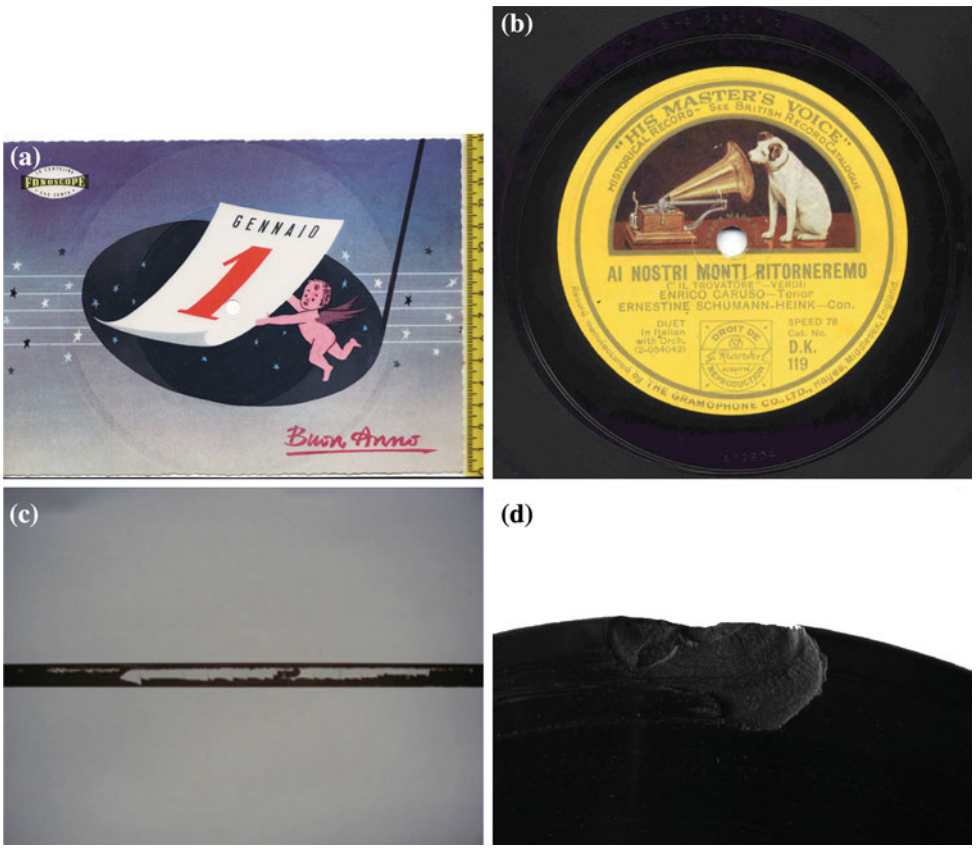

Fig. 3 (**a**) a sound postcard: it looked like a standard postcard on the back, but on the front an analogue recording was engraved in a thin layer of laminate. Sound postcards were usually made by small firms, and the recording quality was extremely low; in this case the importance of storing the picture in with the preservation copy is particularly evident. (**b**) displays a label of His Master's Voice disk: DK 119 (on the label, right) is the catalog number; 2-054042 (on the label, left, and at the top of the mirror) is a second catalog number (as its minor typographic importance, probably it is the first issue catalog 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 catalog number corresponds to a second issue (2), 30 cm diameter (0), Italian catalog (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**, **d**) show two typical corruptions in a tape and in a disk, respectively: this information should be stored with the preservation copy also, in order to have a deep insight the artifacts of the audio signal

acquisition of documents for conservative aims (preservation copies) and the archiving for common use (access copies).

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. 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 [4,19,20].
2. A description of the irregularities in the playback speed of analog recordings (wow and flutter[7]): in disks, a spindle hole not precisely centered and/or the warping of the disk 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

[7] Wow and flutter are audio distortions perceived as an undesired frequency modulation in the range of [10]: (i) wow from 0.5 to 6 Hz, (ii) flutter from 6 to 100 Hz. The distortions are introduced to a signal by an irregular velocity of the analog medium. As the irregularities can originate from various mechanisms, the resulting parasitic frequency modulations can range from periodic to accidental, having different instantaneous values.

Fig. 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, it is possible to locate automatically the imperfections occurred during the A/D transfer (see Sect. 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. 4).

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 320×240 pixels resolution video with medium quality DivX compression yielded satisfactory results.

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. A comprehensive tutorial about audio fingerprinting techniques and applications can be found in [22].

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 [13]. 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.

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

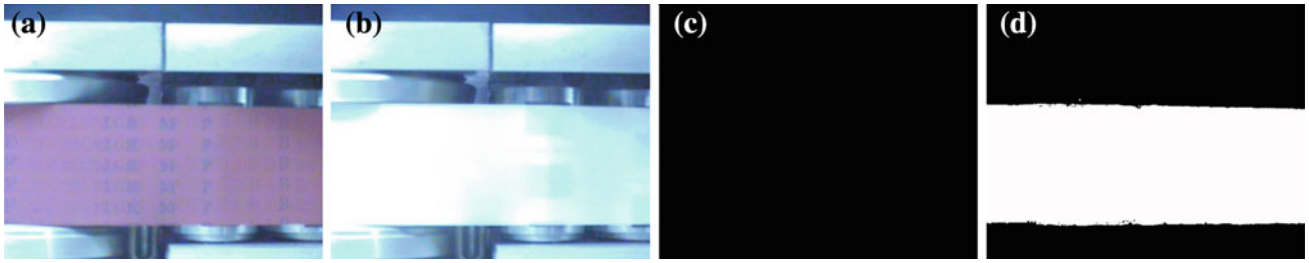

Fig. 5 (**a**, **b**) show source frames from the video of a winding tape, while (**c**, **d**) show the corresponding processed images

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 disk.

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. 5. In this case, we have employed background subtraction with automatic thresholding [62] 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 5c is completely black as no significant changes have been detected between the current frame of Fig. 5a and the background image. In Fig. 5d a major change has occurred (white pixels) in the source frame shown in Fig. 5b (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 (the reader is referred to [62] for implementation details) 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. 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 an 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 [45].

Figure 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. 6b 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.

Figure 6d shows how a tape splice can be detected. The Hough transform [60] 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. 6f a connected component corresponding to the dot in Fig. 6e is detected. The system can, therefore, annotate the corresponding frame linking it to the specific sound event marked by the felt-tip pen sign.

5.2 Warped phonographic disks

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 disks. We have employed a feature tracking algorithm known as the Lucas–Kanade tracker [62]. 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. Figure 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. 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. 8.

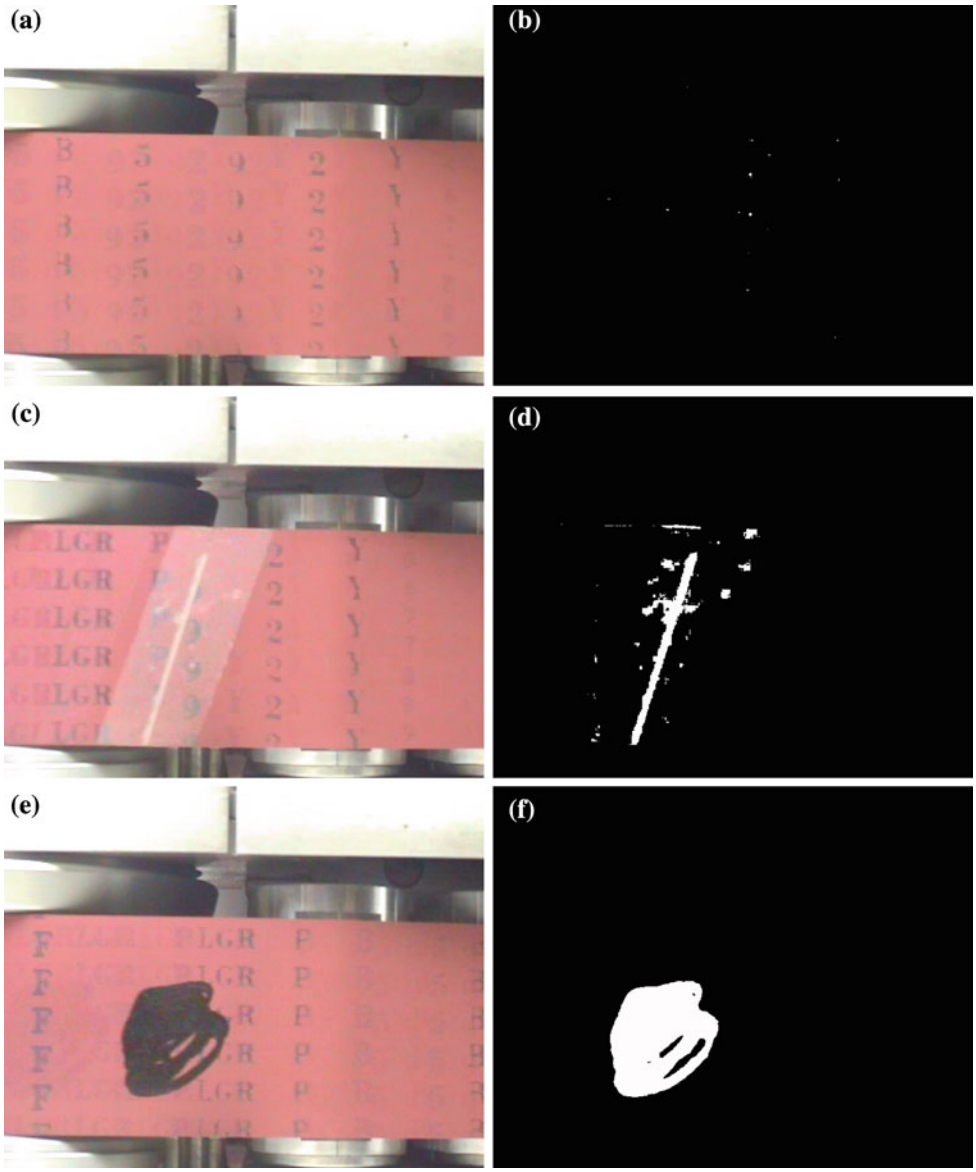

Fig. 6 Automatic discontinuities extraction from a winding tape (splices, marks)

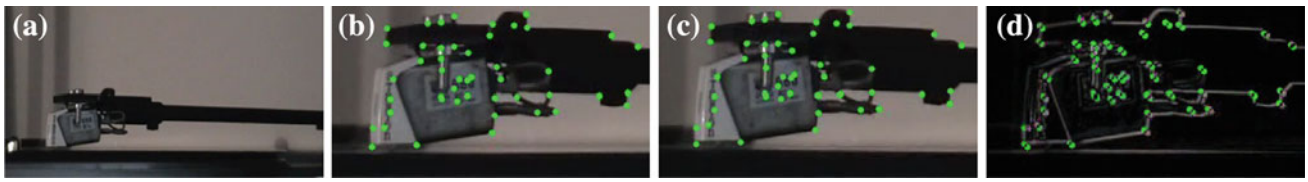

Fig. 7 Processed frames from a video of a oscillating record player's arm. (**a**) Photo of the turntable arm; (**b**) Lowest position of the arm in an oscillation, (**c**) its highest position. (**b**, **c**) show Lucas–Kanade features detected on the arm's head and tracked through the oscillation. (**d**) shows the differences between lowest and highest positions

Figure 8 shows the temporal evolution of the y coordinate of a feature located on the armć6s 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-frame gap between Fig. 7b and c, which is consistent to the period of the oscillations shown in Fig. 8.

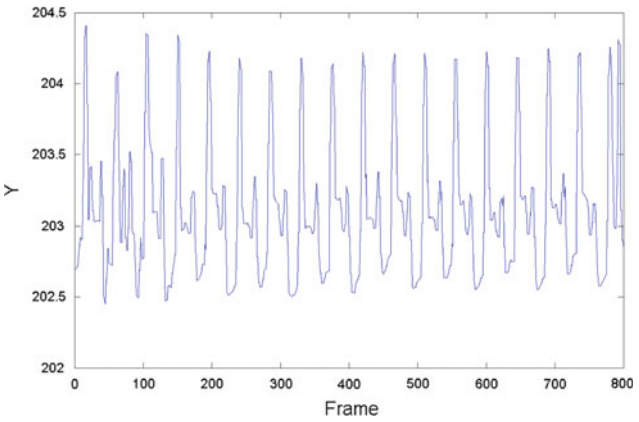

Fig. 8 Temporal evolution of the y coordinate of a Lucas–Kanade feature located on the arm's head. It can be seen clearly how the oscillations indicate a deformed disk

Fig. 9 Disk and spindle hole contours automatically detected via the integrodifferential operator

5.3 Off-centered phonographic disk

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 disk, 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-known warp effect that introduces a pitch variation in the audio signal. To accomplish this automatically we have exploited the consolidated literature on iris detection [45]. 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 [45]. The integrodifferential operator has the following form:

$$\max_{(r,x_0,y_0)} \left| G_\sigma(r) * \frac{\partial}{\partial r} \oint_{r,x_0,y_0} \frac{I(x,y)}{2\pi r} \mathrm{d}s \right| \tag{1}$$

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 σ. In other words, the operator works as circular edge detector and provides the center coordinates and the radius of the strongest circular edge in the image. In our implementation, we extracted the outer contour of the disk first and then rerun the operator on the image for detecting the spindle hole contour as shown in Fig. 9. The second pass can be computed very fast as it takes advantage of the known geometrical properties of vinyl disks. 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. In our setup, the disk was laying on a plane parallel to the image and the spindle hole was on-axis with the camera's optical axis. Although this constraint is not particularly restrictive for a dedicated setup in an audio laboratory, a step further can be taken by removing this assumption and considering perspective deformations given by out-of-axis images as discussed in [25,26].

Having detected the outer boundary of the disk and the spindle hole contour, the calculation of the offset between their centers is trivial. In the experiment reported in Fig. 9, the estimated offset was 1,414 pixels corresponding to 0.22 cm. The processing described in this subsection can be performed on-line in real time. The experiments shown in Figs. 5, 6 and 7, have been carried out on off-line 320 × 240 resolution video sequences with an above real-time frame rate processing performance of 50 frames/s 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 Figs. 5, 6 and 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 setup for audio laboratories and audio digital libraries.

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 [43]. 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 [16]. 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 video capture of the A/D transfer of a warped phonographic disk, 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. 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 [47]. 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 article.

6 Audio data extraction and alignment from phonographic disk

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

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 disk. 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; and (j) they can be used for mass digitization.

In the literature, there are several approaches to this problem [23,31,64]. In general, they are 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) [54] 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; and (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, and turnover). Moreover, PoG allows the user to process the signal by means of several audio restoration algorithms. The software automatically finds the record center 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 [19]. The complete process is depicted in Fig. 10.

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[8] based on the use of a suppression rule, which considers the

[8] Audio restoration algorithms can be divided in three categories [21]:

(a) *frequency-domain* methods, such as various forms of noncasual Wiener filtering or spectral subtraction schemes [12,44,29] and recent algorithms that attempt to incorporate knowledge of the

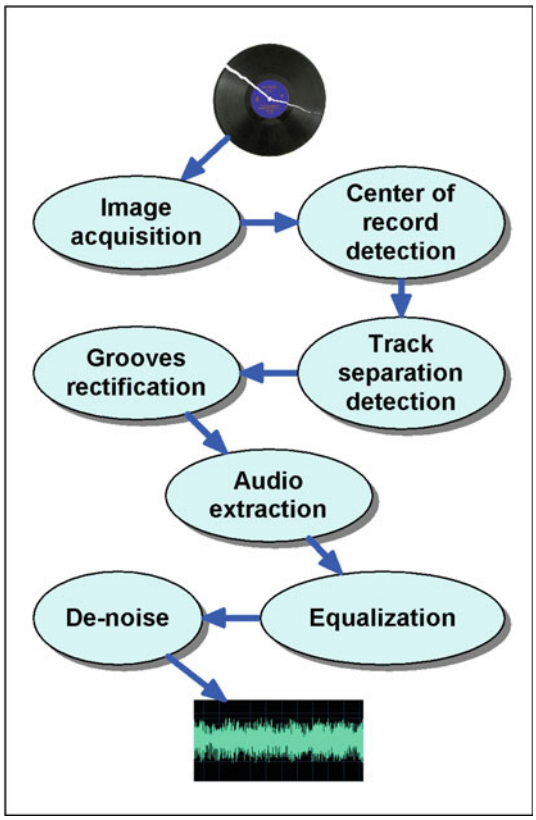

Fig. 10 Photos of GHOSTS schema

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 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 [19], 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.

Photos of Ghosts 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; and (c) the active preservation of carriers with heavy degradation (breakage, flaking, and exudation).

6.2 Audio alignment

The typical application of audio alignment is the comparison of two alternative performances of the same music work. This comparison can be helpful for musicologists to study the style of different conductors and performers, and it can also be exploited to re-synthesize performances adding new expressive parameters. In the case of classical music, alignment can be carried out also between the recording of the performance and a digital representation of the score; yet, audio to audio alignment may be the only option for genres that are not commonly represented by a standard notation, such as ethnic or electro-acoustic music. The alignment of two audio recordings can be a useful tool also when two different versions of the same recording session are to be compared. For instance, in the case of electro-acoustic music, the available recordings of a given work may differ because of different post-processing and editing that have been applied before publication [55]. In this case, alignment allows musicologists to highlight possible cuts and insertions of new material in the recordings, to detect the usage of previously released material inside a new composition, and to compare the temporal and spectral features in corresponding parts also when they have different playback speeds.

We propose to apply alignment techniques to the comparison of alternative digital acquisitions of the same disk. In particular, PoG can be compared to the acquisition based on analog playback. It is likely that the recording speeds differ slightly depending on the technique and that there can be local differences depending on the quality of the equipment. Moreover, the two approaches may give different results in terms of robustness to local damages on the disk surface. For this reason, we propose to use automatic alignment as

Footnote 8 continued
human auditory system [65,66]; 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 [33–35,46,52,56]: 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 [30] is used.

a tool to compare the characteristics of digital acquisitions of a given disk and to evaluate objectively the quality of the proposed technique.

Audio to audio matching is usually based on a preprocessing of the recordings in order to extract relevant features that are able to generalize their main characteristics. A popular descriptor is the chroma-based representations. The basic idea is that all the components of the spectrum are conflated into a single octave, obtaining a particular signature of a polyphonic signal. Alternatively, as presented in [49], audio recordings can be segmented in coherent parts with stable pitch components, and a set of bandpass filters can be computed for each segment around the main peaks in the frequency domain. Once a set of descriptors is extracted from the two audio signals, the global matching can be carried out using dynamic programming approaches to compute the local and global distance between frames in the recording, for instance Dynamic Time Warping (DTW), or statistical modeling of the temporal and spectral differences between the two recordings, for which the most popular tools are Hidden Markov Models (HMMs). Both approaches have been largely exploited in speech recognition [58] and in music identification. In the latter case, a variant of DTW has been proposed in [50] for off-line alignment using chroma features, while a real-time version of DTW has been presented in [27] and an approach to alignment based on HMMs is described in [49].

Both in the case of DTW and HMMs, the global alignment is computed from a local distance using a dynamic programming approach. The main difference is that HMMs require that a model is built from one of the recordings, which becomes the reference signal against which the other recording is compared, while DTW can be carried out directly from the signal parameters without the need of using a particular recording as the reference. Another important difference is that HMMs need to be trained with a number of examples, which might not be available in some application domains, while DTW is simply based on the notion of local distance between audio frames of the two recordings. For these reasons, DTW is proposed to compute the alignment.

The first step in the definition of a distance between two recordings regards the choice of the acoustic parameters that are to be used. Given the relevance of spectral information, the similarity function is normally based on the frequency representation of the signal. To highlight also short local mismatches due to small scratches on the record surface, we choose to use small windows of the signal, of 2048 points with a sampling rate of 44.1 kHz, using a hop size between two subsequent windows of 1024 points. These parameters give a time resolution of the alignment of about 23 ms.

After choosing how to describe the digital recordings, a suitable distance function has to be chosen. Many distances have been proposed in the literature to measure the distance between two spectra, ranging from cross correlation, spectral

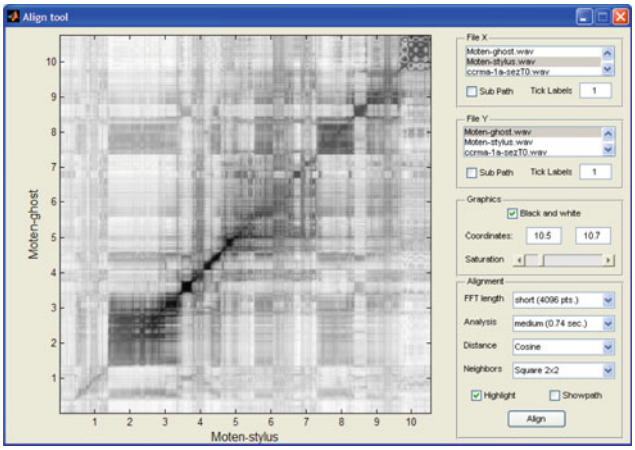

Fig. 11 Visual representation of the similarities between two audio signals. *X*-axis: audio signal extracted by means of turntable; *y*-axis: audio signal generated from a photo of the disk by means of the PoG system (see Sect. 6.1)

flux, to L1 and L2 norms. We propose to use the cosine of the angle between the vectors representing the amplitude of the Fourier transform, which is a well-known measure used typically in information retrieval. Thus, given two recordings f and g, the local distance d(m,n) between two frames can be computed according to equation:

$$d(m,n) = \frac{\sum_{i=1}^{K} F_m(i)\, G_n(i)}{\|F_m\| \;\; \|G_n\|} \tag{2}$$

where Fm (Gn) is the magnitude spectrum of frame m (n) of recording f (g), while in our application $K = 2048$ points. Local distance can be represented by a distance matrix, as shown in Fig. 11, which can be used as a visual representation of the similarities between two recordings. As it can be seen from the Fig. 11, the main similarities are along the diagonal of the matrix, where large dark squares correspond to long sustained notes and brighter areas represent a low degree of similarity between two frames. In practice, the local distance needs to be computed only in proximity of the main diagonal, in order to reduce computational cost.

After the local distance matrix is computed, DTW finds the best aligning path according to equations:

$$c(m,n) = \min \begin{cases} c(m-1,n-1) + 1.5\; d(m,n) \\ c(m-1,n) + d(m,n) \\ c(m,n-1) + d(m,n) \end{cases} \tag{3}$$

$$p(m,n) = \arg\min \begin{cases} c(m-1,n-1) + 1.5\; d(m,n) \\ c(m-1,n) + d(m,n) \\ c(m,n-1) + d(m,n) \end{cases} \tag{4}$$

where $c(m,n)$ is the cumulative distance between the two recordings, computed for each couple of frames. It is possible to compute the global optimal path that starts in point [1,1] and stops in any chosen point through a backtracking procedure that exploits the information stored in $p(m,n)$. It

has to be noted that there have been proposed many different combinations of neighbor points to compute the minimization. The results presented in this article have been computed using this equation, which is based on just three neighbors located on a square.

7 Experimental results

In this section we present our experimental results of applying the above-described techniques related to metadata and audio data extraction, comparing the different signals by means of the audio alignment techniques. We conducted a series of experiments with real usage data from different international audio archives. Examples generated by the methods described in this article are available at: http://avires.dimi.uniud.it/tmp/DL/Experimental_Results.html

7.1 Case study #1: a Chattanooga blues

As first case study, we selected the double-sided 78-rpm shellac disk Okeh 8457—OK 8102 and focused our attention on the song *A Chattanooga Blues*.

The performers are Mary H. Bradford (vocal) with Bennie Moten's Kansas City Orchestra: Lammar Wright, cornet; Thamon Hayes, trombone; Woodie Walder, clarinet; Bennie Moten, piano/leader; George Tall, banjo; and Willie Hall, drum. September 1923. This is an acoustic recording made prior to the use of microphones. Bennie Moten is today remembered as the leader of a band that partly became the nucleus of the original Count Basie Orchestra. He was a fine ragtime-oriented pianist who led the top territory band of the 1920s, an orchestra that really set the standard for Kansas City jazz. Moten formed his group in 1922 and the following year they made their first recordings.

The audio signal was extracted in two ways:

1. By means of the Rek-O-Kut-Rodine 3 turntable; the A/D transfer was carried out with RME Fireface 400 at 44.1 kHz, 16 bit; no equalization curve has been applied.
2. Using PoG system; the image was taken at 4800 dpi, 8 bit grayscale, without digital correction.

Finally, the alignment method described in Sect. 6.2 was used to compare the differences/similarities between these two audio signals. In this way, interesting metadata about the A/D transfer process and the original carrier can be extracted.

7.1.1 Alignment curve

By comparing the two signals, it is possible to point out the discrepancies between the angular velocities used during the disk playing, as shown in Fig. 12 for 15 s of the two audio signals (from 15 to 30 s from the beginning of the recordings). The virtual *velocity* of the PoG system is perfectly constant, and is given by the number of pixels per second read by the software; therefore, the solid curve shows the imperfections of the A/D transfer system, including an imprecise number of RPMs and possible acceleration and deceleration of the turntable during the playing. In our case, the velocity of the audio signal generated by PoG is greater than that extracted with the turntable, despite we set both to 80 rpm (1923 USA Okeh acoustic recording). Moreover, local differences between the two signals are represented by local changes in the slope of the alignment curve. In this way, we have a tool for taking into account some imperfections of the A/D transfer process.

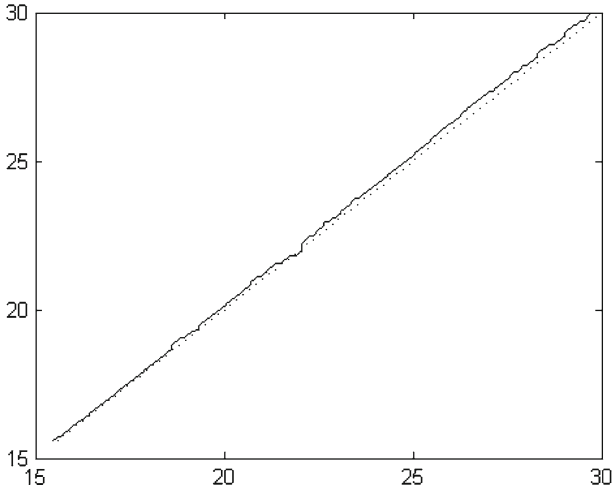

Fig. 12 Case study #1: Alignment curve (*solid line*), in comparison with the bisector (*dashed line*) from 15 to 30 s. *X*-axis: audio signal extracted by means of turntable; *y*-axis: audio signal generated by PoG

7.1.2 Visual representation of the similarities

Figure 13 shows the main similarities between the two signals: brighter areas represent a low degree of similarity between two frames. In the middle of the excerpt there are areas with a low similarity degree: in fact, in this interval the voice recorded in the signal is very distorted. These distortions are *performed* in different manners by the two systems. In this way, we have a tool able to describe serious corruptions of the recording.

7.1.3 Graph of the differences

Figure 14 reports the similarities and the differences between the first 30 s of two signals after alignment, showing that the signal generated by PoG is very different from the re-recording in proximities of local disturbances (scratches and crackles). The local minimum values of the function plotted in Fig. 14 give an estimation of the disk local corruptions.

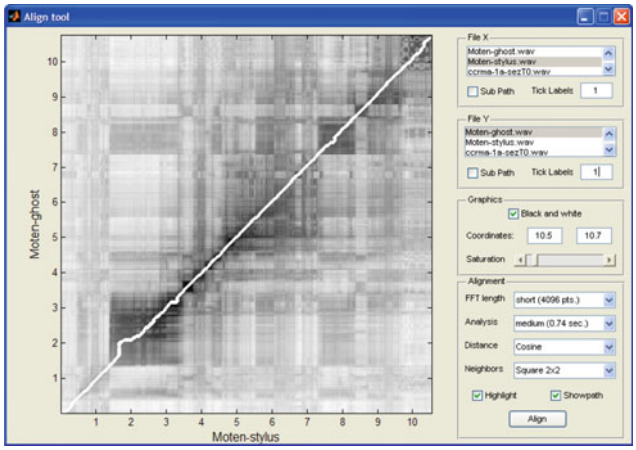

Fig. 13 Case study #1: Visual representation of the similarities between two audio signals. *X*-axis: audio signal extracted by means of turntable; *y*-axis: audio signal generated by PoG. The alignment between the two signals is represented by a *white curve*

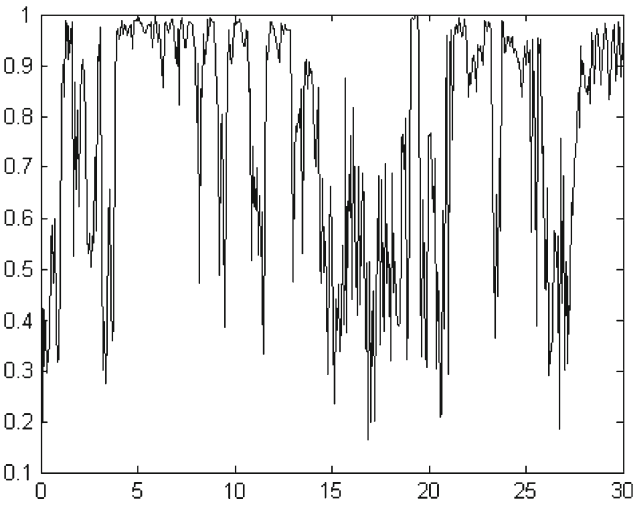

Fig. 14 Case study #1: The graph of the differences between the two audio signals along the alignment curve. *X*-axis: time using PoG signal as a reference; *y*-axis: a similarity degree scaled from 0 to 1

7.2 Case study #2: Imitazione degli animali

The second case study addresses the same example given in Sect. 5. We selected the double-sided 78-rpm shellac disk Piccola Meraviglia 47-148 and focused our attention on the track *Imitazione degli animali, parte I* (Fig. 15). This track contains speech and environmental sounds of animals. The audio signal was extracted in the same conditions and with the same setting described in Sect.7.1.

7.2.1 Alignment curve and visual representation of the similarities

We compare the audio signal extracted by means of turntable with the audio signal generated by PoG. Also in this case,

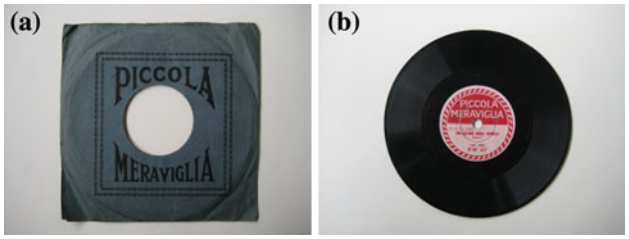

Fig. 15 Case study #2: Cover and disk surface

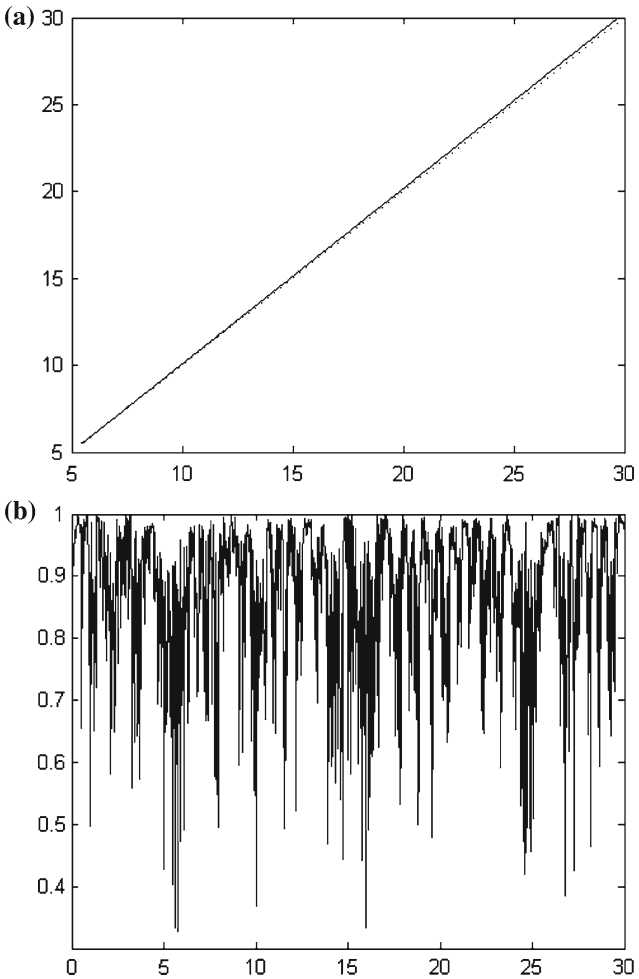

Fig. 16 Case study #2: Comparison between the audio signal extracted by means of turntable and the audio signal generated by PoG. (**a**) Alignment curve (*solid line*), in comparison with the bisector (*dashed line*); (**b**) graph of the differences along time, using PoG signal as the reference

the alignment highlights a difference in the angular velocities, as shown in Fig.16a, while Fig. 16b shows the similarities between the two aligned audio signals. Although also in this case there is a small difference between the RPMs of the turntable and the ones of Pog, the alignment curve is more regular than in Case study #1, probably because the differences are due to a different sensitiveness to very small defects in the grooves. This hypothesis is confirmed also by the plot of the differences between the two signals, which are

more similar in average with local minima that span for few samples. The use of alignment can give an initial assessment of the quality of the A/D transfer.

8 Concluding remarks

The objective of this article 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 study has presented:

1. A novel system able to synthesize the audio signal from a still image of a phonographic disk surface.
2. A software to extract metadata from photos and video shootings of audio carriers.
3. An alignment technique to compare alternative digital acquisitions.
4. Two case studies, in which the alignment tool is used to annotate disk corruptions.

This study summarizes a number of experiences in several research/applied project on Digital Audio Archives and Audio Access, carried out by the authors, 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 article 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"; and "Historic Archive of Contemporary Arts" of the LaBiennale of Venice.

References

1. 3MCompany: High frequency bias requirements for magnetic tape recording. 3M SoundTalk Bull. **1**(2), 1–4 (1968)
2. Adorno, T.W.: Philosophy of New Music. University of Minnesota Press, Minneapolis (2006)
3. AES-11id-2006: AES Information Document for Preservation of Audio Recordings—Extended Term Storage Environment for Multiple Media Archives. AES (2006)
4. AES22-1997: AES Recommended Practice for Audio Preservation and Restoration—Storage and Handling—Storage of Polyester-Base Magnetic Tape. AES (2003)
5. AES28-1997: AES Standard for Audio Preservation and Restoration—Method for Estimating Life Expectancy of Compact Discs (CD-ROM), Based on Effects of Temperature and Relative Humidity (includes Amendment 1-2001). AES (2003)
6. AES31-2-2006: AES standard on Network and File Transfer of Audio—Audio-File Transfer Exchange—File Format for Transferring Digital Audio Data Between Systems of Different Type and Manufacture. AES (2006)
7. AES35-2000 AES Standard for Audio Preservation and Restoration—Method for Estimating Life Expectancy of Magneto-Optical (M-O) Disks, Based on Effects of Temperature and Relative Humidity. AES (2005)
8. AES38-2000: Aes Standard for Audio Preservation and Restoration—Life Expectancy of Information Stored in Recordable Compact Disc Systems—Method for Estimating, Based on Effects of Temperature and Relative Humidity (2005)
9. AES49-2005 AES Standard for Audio Preservation and Restoration—Magnetic Tape—Care and Handling Practices for Extended Usage. AES (2005)
10. A.E. Society: Method for measurement of weighted peak flutter of sound recording and reproducing equipment, AES6-2008. AES Standard (2008)
11. Bertram, H., Cuddihy, E.: Kinetics of the humid aging of magnetic recording tape. IEEE Trans. Magn. **27**, 4388–4395 (1982)
12. Boll, S.: Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoust Speech Signal Process. **ASSSP 27**(2), 113–120 (1979)
13. Boney, L., Tewfik, A., Hamdy, K.: Digital watermarks for audio signals. In: IEEE Proceedings Multimedia pp. 473–480 (1996)
14. Boston, G.: Safeguarding the Documentary Heritage. A Guide to Standards, Recommended Practices and Reference Literature Related to the Preservation of Documents of All Kinds. UNESCO (1988)
15. Brock-Nannestad, G.: The Objective Basis for the Production of High Quality Transfers from Pre-1925 Sound Recordings. In: AES Preprint n °4610 Audio Engineering Society 103rd Convention, pp. 26–29. New York (1997)
16. Brown University Library: Center for digital initiatives (2010). http://pike.services.brown.edu/
17. Burt, L.: Chemical Technology in the Edison Recording Industry. J. Audio Eng. Soc. (**10-11**):712–717 (1977)
18. Calas, M., Fountaine, J. La conservation des documents sonores. CNRS, Paris, France (1996)
19. Canazza, S.: Noise and Representation Systems: A Comparison among Audio Restoration Algorithms. (Lulu Enterprise, USA 2007)
20. Canazza, S., Vidolin, A.: Preserving electroacoustic music. J. New Music Res. **30**(4), 351–363 (2001)
21. Canazza, S., Vidolin, A.: Special issue on preserving electroacoustic music. J. New Music Res. **30**(4) (2001)
22. Cano, P., Batlle, E., Kalker, T., Haitsma, J.: A review of audio fingerprinting. J. VLSI Signal Process. **41**, 271–284 (2005)
23. Cavaglieri, S., Johnsen, O., Bapst, F.: Optical retrieval and storage of analog sound recordings. In: AES (ed.) Proceedings of AES 20th International Conference. Budapest, Hungary (2001)
24. Cohen, E.: Preservation of audio in folk heritage collections in crisis. In: Proceedings of Council on Library and Information Resources. Washington, DC, USA (2001)
25. Daugman, J.: How iris recognition works. IEEE Transactions on Circuits and Systems for Video Technology **14**(1), 21–30 (2004)

26. Daugman, J.: New methods in iris recognition. IEEE Trans Syst Man Cybern B Cybern **37**(5), 1167–1175 (2007)
27. Dixon, S., Widmer, G.: Match: a music alignment tool chest. In: Proceedings of the International Conference of Music Information Retrieval, pp. 492–497 (2005)
28. EBU: Specification of the Broadcast Wave Format: A Format for Audio Data Files in Broadcasting—Tech 3285. EBU (1997)
29. Ephraim, Y., Malah, D.: Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator. IEEE Trans. Acoust. Speech Signal Process. **32**(6), 1109–1121
30. Esquef, P.A.A., Valimaki, V., Karjalainen, M.: Restoration and enhancement of solo guitar recordings based on sound source modeling. J. Audio Eng. Soc. **50**(4), 227–236 (2002)
31. Fedeyev, V., Haber, C.: Reconstruction of mechanically recorded sound by image processing. J. Audio Eng. Soc. **51**(12), 1172–1185 (2003)
32. Gibson, G.: Magnetic tape deterioration: recognition, recovery and prevention (1996). http://www.unesco.org/webworld/ramp/html/r9704e/r9704e11.htm
33. Grancharov, V., Samuelsson, J., Kleijn, B.: Noise-dependent postfiltering. Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP) **1**, 457–460 (2004)
34. Grancharov, V., Samuelsson, J., Kleijn, B.: Improved Kalman filtering for speech enhancement. In: Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP) **1**, 1109–1112 (2005)
35. Grancharov, V., Samuelsson, J., Kleijn, B.: On casual algorithms for speech enhancement. Trans. Audio Speech Lang. Process. **14**(3), 273–276 (2006)
36. Hart, M.: Preserving our musical heritage: a musician's outreach to audio engineers. J. Audio Eng. Soc. **49**(7–8) (2001)
37. IASA-TC 03: The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy. IASA Technical Committee (2005)
38. IASA-TC 04: Guidelines on the Production and Preservation of Digital Objects. IASA Technical Committee (2004)
39. IFLA/UNESCO: Safeguarding our Documentary Heritage/Conservation préventive du patrimoine documentaire/Salvaguardando nuestro patrimonio documental. CD-ROM Bi-lingual: English/French/Spanish. UNESCO "Memory of the World" Programme, French Ministry of Culture and Communication (2000)
40. Khanna, S.: Vinyl compound for the phonographic industry. J. Audio Eng. Soc. (**10–11**), 712–717 (1977)
41. Knight, G.: Factors relating to long term storage of magnetic tape. Phonograph. Bull. (**18**), 16–37 (1977)
42. Laurent, S.: *The Care of Cylinders and Discs*. (Technical Coordinating. Committee, Milton Keynes 1997)
43. Library of Congress: Metadata encoding and transmission standard (METS) (2010). http://www.loc.gov/standards/mets/
44. Lim, J., Oppenheim, A.: All-pole modeling of degraded speech. IEEE Trans. Acoust. Speech Signal Process. **26**(3), 197–210 (1978)
45. Liu, Y., Zhang, D., Lu, G., Ma, W.: A survey of content-based image retrieval with high-level semantics. Pattern Recognit. **40**(1), 262–282 (2007)
46. Ma, N., Bouchard, M., Goubran, R.A.: Speech enhancement using a masking threshold constrained Kalman filter and its heuristic implementations. IEEE Trans. Speech Audio Lang. Process. **14**(1), 19–32 (2006)
47. Manjunath, B., Salembier, P., Sikora, T.: Introduction to MPEG-7: Multimedia Content Description Interface. Wiley, New York (2002)
48. Miller, D.: The Science of Musical Sounds. Macmillan, New York (1922)
49. Miotto, R., Orio, N.: Automatic identification of music works through audio matching. In: Proceedings of 11th European Conference on Digital Libraries, pp. 124–135 (2007)
50. Müller, M., Kurth, F., Clausen, F.: Audio matching via chroma-based statistical features. In: Proceedings of the International Conference of Music Information Retrieval, pp. 288–295 (2005)
51. Negroponte, N.: Being Digital. Vintage Books, New York (1995)
52. Niedźwiecki, M., Cisowski, K.: Adaptive scheme for elimination of broadband noise and impulsive disturbances from AR and ARMA signals. IEEE Trans. Signal Process. **44**(3), 967–982 (1996)
53. Orcalli, A.: On the methodologies of audio restoration. J. New Music Res. **30**(4), 307–322 (2001)
54. Orio, N., Snidaro, L., Canazza, S.: Semi-automatic metadata extraction from shellac and vinyl disc. In: Proceedings of Workshop on Digital Preservation Weaving Factory for Analogue Audio Collections. Firenze University Press, Firenze, Italy, pp. 38–45 (2008)
55. Orio, N., Zattra, L.: Audio matching for the philological analysis of electro-acoustic music. In: Proceedings of the International Computer Music Conference, pp. 157–164 (2007)
56. Paliwal, K., Basu, A.: A speech enhancement method based on Kalman filtering. Proc. IEEE Int. Conf. Acoust. Speech Signal Audio Process. vol. **12**, pp. 177–180 (1987)
57. Pickett, A., Lemcoe, M.: Preservation and Storage of Sound Recordings. ARSC, Washington, DC, USA (1991)
58. Rabiner, L., Juang, B.: Fundamentals of Speech Recognition. Prentice-Hall, Englewood Cliffs, NJ (1993)
59. Schüller, D.: The ethics of preservation, restoration, and re-issues of historical sound recordings. J. Audio Eng. Soc. **39**(12), 1014–1016 (1991)
60. Shapiro, L., Stockman, G.: Computer Vision. Prentice-Hall, Upper Saddle River (2001)
61. Smith, A.: Why digitize? In: Proceedings of Council on Library and Information Resources. Washington, DC, USA (1999)
62. Snidaro, L., Foresti, G.L.: Real-time thresholding with Euler numbers. Pattern Recognit. Lett. **24**(9-10), 1533–1544 (2003)
63. Storm, W.: The establishment of international re-recording standards. Phonograph. Bull. **27**, 5–12 (1980)
64. Stotzer, S., Johnsen, O., Bapst, F., Sudan, C., Ingol, R.: Phonographic sound extraction using image and signal processing. In: Proc. ICASSP, **4**, 289–292 (2004)
65. Tsoukalas, D., Mourjopoulos, J., Kokkinakis, G.: Speech enhancement based on audible noise suppression. IEEE Trans. Acoust. Speech Signal Process. **5**(6), 497–514 (1997)
66. Virag, N.: Single channel speech enhancement based on masking properties of the human auditory system. IEEE Trans. Acoust. Speech Signal Process. **7**(2), 126–137 (1999)