UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Head Office: Università degli Studi di Padova

Department of Information Engineering (DEI)

PH.D. COURSE IN: Information Engineering

CURRICULUM: Information Communication Technology (ICT)

SERIES: XXXI

CULTURAL CONTEXT-AWARE MODELS AND IT APPLICATIONS FOR THE EXPLOITATION OF MUSICAL HERITAGE

Coordinator: Ch.mo Prof. Andrea Neviani

Supervisor: Ch.mo Prof. Sergio Canazza

Ph.D. Student: Niccolò Pretto

Abstract

Information engineering has always expanded its scope by inspiring innovation in different scientific disciplines. In particular, in the last sixty years, music and engineering have forged a strong connection in the discipline known as "Sound and Music Computing". Musical heritage is a paradigmatic case that includes several multi-faceted cultural artifacts and traditions. Several issues arise from the analog-digital transfer of cultural objects, concerning their creation, preservation, access, analysis and experiencing. The keystone is the relationship of these digitized cultural objects with their carrier and cultural context. The terms "cultural context" and "cultural context awareness" are delineated, alongside the concepts of contextual information and metadata. Since they maintain the integrity of the object, its meaning and cultural context, their role is critical. This thesis explores three main case studies concerning historical audio recordings and ancient musical instruments, aiming to delineate models to preserve, analyze, access and experience the digital versions of these three prominent examples of musical heritage.

The first case study concerns analog magnetic tapes, and, in particular, tape music, a particular experimental music born in the second half of the XX century. This case study has relevant implications from the musicology, philology and archivists' points of view, since the carrier has a paramount role and the tight connection with its content can easily break during the digitization process or the access phase. With the aim to help musicologists and audio technicians in their work, several tools based on Artificial Intelligence are evaluated in tasks such as the discontinuity detection and equalization recognition. By considering the peculiarities of tape music, the philological problem of stemmatics in digitized audio documents is tackled: an algorithm based on phylogenetic techniques is proposed and assessed, confirming the suitability of these techniques for this task. Then, a methodology for a historically faithful access to digitized tape music recordings is introduced, by considering contextual information and its relationship with the carrier and the replay device. Based on this methodology, an Android app which virtualizes a tape recorder is presented, together with its assessment. Furthermore, two web applications are proposed to faithfully experience digitized 78 rpm discs and magnetic tape recordings, respectively. Finally, a prototype of web application for musicological analysis is presented. This aims to concentrate relevant part of the knowledge acquired in this work into a single interface.

The second case study is a corpus of Arab-Andalusian music, suitable for computational research, which opens new opportunities to musicological studies by applying data-driven analysis. The description of the corpus is based on the five criteria formalized in the CompMusic project of the University Pompeu Fabra of Barcelona: purpose, coverage, completeness, quality and re-usability. Four Jupyter notebooks were developed with the aim to provide a useful tool

for computational musicologists for analyzing and using data and metadata of such corpus.

The third case study concerns an exceptional historical musical instrument: an ancient Pan flute exhibited at the Museum of Archaeological Sciences and Art of the University of Padova. The final objective was the creation of a multimedia installation to valorize this precious artifact and to allow visitors to interact with the archaeological find and to learn its history. The case study provided the opportunity to study a methodology suitable for the valorization of this ancient musical instrument, but also extendible to other artifacts or museum collections. Both the methodology and the resulting multimedia installation are presented, followed by the assessment carried out by a multidisciplinary group of experts.

Sommario

L'ingegneria dell'informazione ha sempre abilitato innovazione in moltissimi settori scientifici. In particolare, negli ultimi sessant'anni, la ricerca nel campo della *Sound and Music Computing*, ha coniugato musica e ingegneria in un'unica disciplina. Il patrimonio musicale è un caso paradigmatico che include numerose tipologie di reperti e di tradizioni. Il trasferimento nel dominio digitale di oggetti culturali implica diversi problemi legati alla loro creazione, conservazione, accesso, analisi e fruizione. Il cardine è la relazione degli oggetti culturali digitalizzati con il loro supporto originario e con il loro contesto culturale. In questa tesi, i concetti di *cultural context* e *cultural context awareness* sono descritti in relazione ai metadati e alle informazioni contestuali necessarie per mantenere l'integrità, il significato e il contesto culturale dell'oggetto digitalizzato. Questo lavoro approfondisce tre studi di caso riguardanti i documenti sonori storici e gli strumenti musicali antichi, con l'obiettivo di formalizzare dei modelli per la conservazione, l'analisi, l'accesso e la fruizione del patrimonio musicale.

Il primo studio di caso riguarda le registrazioni su nastro magnetico di *tape music*, una particolare musica sperimentale della seconda metà del Novecento, che ha rilevanti implicazioni dal punto di vista musicologico, filologico e archivistico. In questo caso, il supporto ha un ruolo di straordinaria importanza e una stretta relazione con il suo contenuto, che può facilmente perdersi durante la digitalizzazione o la fruizione. È stato sviluppato un insieme di strumenti automatici basati su tecniche di intelligenza artificiale per supportare musicologi e tecnici audio nel loro lavoro. È quindi stato studiato il problema filologico relativo alla creazione dello *stemma codicum* nei documenti sonori digitalizzati, ben considerando le peculiarità della *tape music*. A tale scopo, è stato sviluppato un algoritmo basato su tecniche di filogenetica, successivamente testato con ottimi risultati. A seguire, la tesi definisce una innovativa metodologia per una fruizione storicamente fedele dei documenti sonori digitalizzati, in grado di valorizzare le informazioni contestuali, il supporto originale e il suo dispositivo di riproduzione. Sulla base di questa, è stata sviluppata un'app Android che virtualizza un magnetofono, presentata assieme alla valutazione di un gruppo multidisciplinare di esperti. Inoltre, sono presentate due applicazioni che propongono la virtualizzazione rispettivamente di un grammofono per dischi a 78 giri e un magnetofono. La tesi propone anche una interfaccia web per l'analisi musicologica, che cerca di racchiudere in un'unica applicazione gli strumenti e le conoscenze acquisite.

Un secondo studio di caso è basato su di un corpus di musica Araba-Andalusa, creato appositamente per soddisfare le esigenze della *computational musicology*, e finalizzato a uno studio *data-driven* di questa importante tradizione musicale. Il *corpus* è presentato sulla base di cinque criteri formalizzati nel progetto CompMusic dell'Università Pompeu Fabra di Barcellona: scopo,

copertura, completezza, qualità e riusabilità. Sono stati sviluppati quattro notebook Jupyter al fine di fornire uno strumento per l'utilizzo e l'analisi del corpus.

Il terzo studio di caso concerne un eccezionale strumento musicale antico: un flauto di Pan esposto al Museo di Scienze Archeologiche e d'Arte dell'Università di Padova. È stata progettata e realizzata un'installazione multimediale per la valorizzazione di questo straordinario reperto, in grado di fornire ai visitatori del museo la possibilità di interagire con la sua copia digitale e la sua storia. Questo studio ha permesso di approfondire e formalizzare una metodologia per la valorizzazione di questo strumento musicale antico, che può essere estesa ad altri reperti o collezioni museali. Sia la metodologia, sia l'installazione sono state valutate da un gruppo multidisciplinare di esperti.

Acknowledgements

Several people deserve my acknowledgements, but first I have to say "Thank you!" to my parents, Annalisa and Antonio: if I can defend my PhD thesis is due to their everyday support. I also deeply thank the other members of my family, Emma, Giulia, Luca, Matteo and Viola for being at my side when I need. Then, I have to say a great "Thank you!" to Valentina, which I met at the beginning of this Ph.D. and, since then, no matter how far away we are, is supporting me every single day without tiring. A big thanks can't miss to my friends Andreo, Balza, Boa, Borgo, Cecky, Chiara, Giulia, Gabe, Hubert, Tata, Treno, Samy, Scrich, Stefy, and, last but not least, Za: with your company (and many drinks), you wonderfully supported me in these hard three (actually more than ten) years.

Huge thanks goes to the person that have more believed in my Ph.D: my supervisor, prof. Sergio Canazza, who bet on me since my master thesis. A deserved thanks is also for all the people that I had the pleasure to meet (and work with) at CSC and let me discover the importance of teamwork for doing research (or just only for having a good break): prof. Federico Avanzini, Federico Altieri, Roberto Barumerli, Matteo Breschigliaro, Federica Bressan, Francesco Carbone, prof. Giovanni De Poli, prof. Carlo Fantozzi, Michele Geronazzo, Riccardo Luise, Maria Maistro, Marcella Mandanici, Dario Marinello, Edoardo Micheloni, Antonio Origlia, Davide Pagin, prof. Antonio Rodà, Alessandro Russo, Simone Spagnol and the "maestro" Alvise Vidolin.

This Ph.D. gave me the great possibility to travel, and the six months in Barcelona will remain as one of the most beautiful experiences of my life, from the professional and the personal points of view. Therefore, first I have to thank prof. Xavier Serra, who gave me the chance to work in his great group: at MTG I met a lot of lovely people and excellent researchers, but I have to acknowledge especially the people that mainly helped me in this Spanish adventure. They are prof. Baris Bozkurt, Rafael Caro Repetto, Emir Demirel, Vsevolod Eremenko, Xavier Favory, Rong Gong, Alastair Porter and Orio Romani. However, Barcelona was not only work, but also great weekends and everyday life! So thanks to my flat mate Judy Kao and all the guys of the UPF4BEER :D of which deserve to be mentioned Federico Franzoni, Federico Mattion, Maria Rauschenberge, Rasoul Nikbakht Silab, Olga Slizovskaia, and Javier Vázquez Corral.

One of the most beautiful part of my Ph.D. was the possibility to collaborate with a lot of great researchers of other fields. So, I have to mention at least Giulia Deotto, Silvia Gasparotto, Alessandra Menegazzi, prof. Simone Milani, Sebastiano Verde and prof. Paola Zanovello. During these three years, I have supervised 22 bachelor or master thesis, so important thanks go to the army of students that helped me to code and develop some of the applications described in this thesis.

Pubblications

Published journal paper

- [36] S.Canazza, C.Fantozzi, and **N.Pretto**. Accessing tape music documents on mobile devices. *ACM Transactions Multimedia Computing Communications and Applications (TOMM)*, 12 (1s):20:1–20:20, October 2015. ISSN 1551-6857. doi: 10.1145/2808200.

- [67] C. Fantozzi, F. Bressan, **N. Pretto**, and S. Canazza. Tape music archives: from preservation to access. *International Journal on Digital Libraries*, 18(3):233–249, September 2017. ISSN 1432-1300. doi: 10.1007/s00799-017-0208-8.

- [144] S. Verde, **N. Pretto**, S. Milani, and S. Canazza. Stay true to the sound of history: Philology, phylogenetics and information engineering in musicology. *Applied Sciences*, 8(2), 2018. ISSN 2076-3417. doi: 10.3390/app8020226.

Accepted journal paper

- [117] **N. Pretto**, C. Fantozzi, E. Micheloni, V. Burini, and S. Canazza. Computing Methodologies Supporting Preservation of Electroacoustic Music from Analog Magnetic Tape. *Computer Music Journal* (accepted).

Accepted book chapter

- [29] V. Burini, S. Canazza, and **N. Pretto**. Archivi di musica elettroacustica: dalla conservazione all'accesso. In *I documenti Sonori. Voce, suono, musica in archivi e raccolte*. Archivi e biblioteche in Piemonte - Direzione Promozione della Cultura, Turismo e Sport della Regione Piemonte, Torino, Italy, 2018.

International and national conference paper

- [116] **N. Pretto** and S. Canazza. Rewind: Simulazione di un'esperienza d'ascolto storicamente fedele di dischi fonografici digitalizzati. *In Proceedings of the XX Colloquium of Musical Informatics, (CIM'14)*, pages 29–34, Roma, Italy, October 2014.

- [8] F. Avanzini, S. Canazza, G. De Poli, C. Fantozzi, **N. Pretto**, A. Rodà, I. Angelini, C. Bettineschi, G. Deotto, E. Faresin, A. Menegazzi, G. Molin, G. Salemi, and P. Zanovello. Archaelogy and virtual acoustics. a pan flute from ancient egypt. *In Proceedings of 12th International Sound and Music Computing Conference (SMC'15)*, pages 31–36, Maynooth, Ireland, August 2015. doi: 10.5281/zenodo.851067.

- [9] F. Avanzini, S. Canazza, G. De Poli, C. Fantozzi, E. Micheloni, **N. Pretto**, A. Rodà, S. Gasparotto, and G. Salemi. Virtual reconstruction of an ancient greek pan flute. *In Proceedings of the 13th International Sound and Music Computing Conference (SMC'16)*, Hamburg, Germany, September 2016. doi: 10.5281/zenodo.851179.

- [101] E. Micheloni, **N. Pretto**, F. Avanzini, S. Canazza, and A. Rodà. Installazioni interattive per la valorizzazione di strumenti musicali antichi: il flauto di Pan del Museo di scienze archeologiche e d'arte dell'Università degli studi di padova. *In Extending interactivity. Atti del XXI CIM - Colloquio di Informatica Musicale (CIM'16)*, pages 203–208, Cagliari, Italy, September 2016.

- [102] E. Micheloni, **N. Pretto**, and S. Canazza. A step toward ai tools for quality control and musicological analysis of digitized analogue recordings: recognition of audio tape equalizations. *In Proceedings of the 11th International Workshop on Artificial Intelligence for Cultural Heritage (AI*CH'17)*, pages 17–24, Bari, Italy, November 2017. URL www.scopus.com/inward/record.uri?eid=2-s2.0-85041462080&partnerID=40&md5=0fabe08bda617c677dfa785145504825.

- [118] **N. Pretto**, B. Bozkurt, R. Caro Repetto, and X. Serra. Nawba recognition for arab-andalusian music using templates from music scores. *In Proceedings of 15th Sound and Music Computing Conference (SMC'18)*, pages 405–410, Limassol, Cyprus, July 2018. doi: 10.5281/zenodo. 1257388.

- [37] R. Caro Repetto, **N. Pretto**, A. Chaachoo, B. Bozkurt, and X. Serra. An open corpus for the computational research of arab-andalusian music. *In Proceedings of the 5th International Conference on Digital Libraries for Musicology (DLfM '18)*, pages 78–86, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-6522-2. doi: 10.1145/3273024.3273025.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Musical Cultural Heritage in the Digital Domain

Nowadays, information engineering has greatly expanded its original scope by inspiring innovation in different scientific disciplines. Information engineering does not provide just tools, but paradigms that can enrich or in some cases revolutionize other disciplines. For example, in linguistics, the view of languages as sets of strings generated by grammars or recognized by computational machines has provided a new foundation to linguistics (e.g. Chomsky hierarchy [43]). The disruptive power of digital technology has been changing drastically the way researchers do their work. In some cases, it has even driven the creation of new derived disciplines such as *digital humanities* [17]. The strength of such multidisciplinary studies is stating the obvious. Human-Computer Interaction (HCI), for example, lies at the intersection between social and behavioral science, and information technology. Nowadays, it has a significant impact on society, economics, culture, etc. [38].

Over 60 years, music and technology have forged such a strong connection that all aspects of the process, from production to distribution and consumption, have become digital. Sound and Music Computing (SMC) is a multi-faceted discipline that led this transformation: SMC research "approaches the whole sound and music communication chain from a multidisciplinary point of view. By combining scientific, technological and artistic methodologies it aims at understanding, modeling and generating sound and music through computational approaches." [16]. This thesis belong to this multidisciplinary branch of information engineering research.

By nature, music is intangible and therefore its heritage is part of *intangible cultural heritage*. UNESCO defines the intangible cultural heritage as "the practices, representations, expressions, knowledge, skills – as well as the instruments, objects, artifacts and cultural spaces associated therewith – that communities, groups and, in some cases, individuals recognize as part of their cultural heritage" [141]. Considering the aforementioned definition, the objects and carriers related to music performance can be considered as part of this intangible cultural heritage, e.g. scores, musical instruments, costumes, posters, audio recordings, etc.

Nevertheless, the concept of intangible is not straightforward by nature and difficult to formalize. As a result, regulations concerning the preservation of musical heritage are fragmented

in several categorizations that often do not consider its peculiarities. For example, the Italian law *Codice dei beni culturali e del paesaggio* (d.lgs. 42/2004 updated in May 7, 2016)[1] gathers geographic maps and music scores in a single category, denoting a lack of a single categorization for musical heritage. Vice versa, audio recordings are usually considered as documentary heritage [82]. The same can be said about ancient musical instruments in museums which are often considered and disciplined as artifacts of completely different nature. These statements do not want to be a critic, but only an example to highlight the vastness and the diversity of this important part of the cultural heritage, and the problems that this heterogeneity implicates.

A further issue derives from the digitization of this heritage for its preservation, access or analysis. Nowadays, dematerialization of artifacts and their reduction to electronic information represents a topical issue [97]. As stated in [17], "to mediate a cultural object, a digital or computational device requires that this object be translated into the digital code that it can understand". This problem is shared by several research areas such as *music information retrieval* (MIR) and *computational musicology* [13], where generating a useful digital representation is vital in order to perform musicologically relevant studies [129]. Information engineering has a paramount role in this process, but the cultural context has to be considered in order to avoid loss of information or even the creation of false and incorrect digital copies. Projects such as CompMusic[2] have emphasized the need to develop engineering methods that are musicologically informed [129]. Nevertheless, several aspects have to be tackled yet.

This work deals with several issues related to three main case studies concerning historical audio recordings and ancient musical instruments, in order to delineate methodologies to preserve, access and experience the digital versions of these prominent examples of musical heritage. Maintaining the relationship of these digitized cultural objects with their cultural context is fundamental. The next section introduces this issue and presents the case studies.

---

[1]"*Sono altresì beni culturali [...] le carte geografiche e gli spartiti musicali aventi carattere di rarità e di pregio*", that can be translated as "rare and valuable geographic maps and music scores are considered as part of the cultural heritage"

[2]`compmusic.upf.edu/` (Retrieved September 18, 2018)

1.2 Toward a model of cultural context

The Oxford dictionary states the term *context* as "the circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood"[3]. The definition of "cultural context" is not unique, but varies on the base of the research field in which is considered. Social science, for example, uses this term to refer the context in which a person lives (often referred as socio-cultural context). Computer science does not explicitly define the cultural context, but delimits the context-awareness as the capability of perceiving the user situation in its main aspects, and to adapt the presentation of the document content accordingly [40].

With the term "cultural context awareness", this work refers to something different. While the aforementioned definitions proposes models and methodologies focused on the people, the user and its devices, the cultural context is referred in this work to an object which can't lose its meaning in its digitized form. Extending the dictionary definition, a digitized object needs its cultural context in order to be "fully understood". This reflects a complementary point of view, which aims to preserve and maintain the cultural object without changing its nature.

This work does not presume to define a new field, but it investigates some case studies in order to explore how the integrity of a digitized object can be preserved by considering the digitization, as well as the access, consumption, experience related to a cultural object. Furthermore, this work proposes a way to exploit these digital cultural objects with the aim to present, valorize or analyze them in their digital domain.

The keystone appears to be the contextual information and metadata, which maintain the integrity of an object, its meaning and cultural context. As stated in [129], meta-information accompanying the content of a digital object can be as important as the actual content itself. Several definitions of metadata can be found in [27], but all of them agree on the main function: to describe the data. In this paper, Bulterman presents its personal definition of metadata as "optional structured descriptions that are publicly available to explicitly assist in locating objects". Furthermore, he highlights the term "optional" stating that "if the descriptions aren't optional, then they are data". From this statement, it is easier to discern "contextual information" from metadata. They are ancillary content-independent information (and so data) which have to be saved in order to maintain the integrity of a document.

This work deals with this kind of information, studying and using them in two different fields which are presented in the next sections: analog audio recordings and historical musical instruments.

1.2.1 Audio Recordings

According to the UNESCO classification [82] of the documentary heritage, audio documents can be divided considering the typology of their carriers in mechanical, magnetic and optical carriers. The mechanical carrier category includes phonograph cylinders, shellac discs, instantaneous discs and microgroove discs. The magnetic carriers typology includes three main types of magnetic tape housing – open spool tapes, cassette tapes and tape cartridges – and magnetic

---

[3]`en.oxforddictionaries.com/definition/context` (Retrieved September 27, 2018)

discs (such as hard discs and floppy discs). The optical carriers mainly include CDs, DVDs, Blue Rays and MiniDiscs. All these media share the materiality, unlike the electronic documents generically classified in [82]. Some magnetic and optical carriers are used to store digital information such as tape cartridges and magnetic discs. Although they share common peculiarities with analog carriers, this work only considers the latter ones, unless otherwise specified.

Analog carriers represents huge part of the musical cultural heritage. Since their introduction (late XIX century), audio recordings have provided unmediated documentation about cultural identity, social and artistic activities and soundscapes ecology [2]. In [127] Dietrich Schüller estimates that the world's stock of audio recordings on magnetic tapes consists of more than 50 millions of hours of materials. Today, audio recordings remain a priceless source of information for a number of research areas such as linguistics, anthropology, and musicology. For example, reel-to-reel tape recorders were the main recording format in professional recording studios until the late 1980s, when digital audio recording techniques disrupted the music world. Inexpensive reel-to-reel tape recorders were also widely used for voice recording at home, and for linguistics and anthropology research activities [67]. Historically, the musical artifact *par excellence* is the musical score, while audio recordings are not yet valorized as they deserve. However, in several fields, audio recordings support leading figures in their work. It is the case of the folklorist Alan Lomax [48, 137]. Its greatest legacy is the preservation and the publication of numerous field recordings of musicians in many folk and blues traditions around the US and Europe. Around 17,400 of Lomax's recordings are freely available online [4]. Roberto Leydi, one of the main Italian ethnomusicologist [130], who published significant studies on Italian folk and popular music [91, 92, 93], before its death, donated 6,000 recordings, 1,000 tapes, 700 musical instruments and more than 10,000 books to Center for Dialectology and Ethnography in Bellinzona (Switzerland).

Several archives have been recently discovered or donated to foundations or universities, and several others will be discovered in the next decades. However, unlike traditional cultural heritages such as paintings or sculptures, audio recordings have a short life expectancy of audio media (years or decades). This is mostly caused by the physical degradation, not only of the carrier, but also of its replay device. Other lesser-known causes endangered this incredible heritage, such as the lack of technical skills [55], lack of good standards or an incorrect digitization process. In the following subsection, an in-depth analysis on preservation methodology is presented.

Preservation of Analogue Audio Documents

Although the physical degradation of an audio carrier can be slowed down through correct preservation policies, it cannot be stopped [23]. Each type of audio carrier degrades in time according to its chemical proprieties [67]. The causes and the syndromes are well-known [105, 79], but the study area concerning degradation mechanisms and evaluation and diagnostic techniques is still an open research field [67], and several challenges in archiving and restoration are not tackled yet [120].

Preservation of analogue audio recordings can be divided in two main fields: *passive* (also re-

---

[4] `research.culturalequity.org/home-audio.jsp` (Retrieved September 27, 2018)

ferred as preservative) and *active preservation*. The first type of preservation can also be splitted in [23]:

- *indirect*, concerning the preservation environment in order to maintain the artifact in the best conditions as possible, for example, by imposing proper storage conditions or handling procedures;

- *direct*, related to the direct intervention on the artifact, such as a treatment in order to stabilize the physical condition of the document without altering its structure and composition.

On the other hand, active preservation involves the data transfer onto new media (rerecording [126]), to prevent an irreversible partial or complete loss of information due to the degradation of the original carrier. The survival of the document information is only possible by renouncing to its materiality, through a constant transfer of the information onto new carriers. Nowadays, it is taken for granted that the new carrier is digital, but the debate on the use of digital technologies was still opened until the early 2000s. In the 80s and 90s, the digitization was considered a method to provide access to an audio document, and not as a mean to preserve it [23]. When the community accepted the concepts "preserve the content, not the carrier" and "distribution is preservation" [47], a new research field was opened concerning the correct preservation, manage and access of the analogue audio document in the digital world. Nevertheless, preservation does not coincide only with the digitization of the audio content, as is often thought: it is an intellectual and practical process, which extends from the preservation to the access and consumption of the audio documents.

The digitization is not a neutral process and has important implications from the philological point of view. During this process, the history of the transmission of the document may be violated, and the documentary unity can be broken, with the result that the digital copy of the tape or record is not trustable in terms of authenticity [67]. In the 1980s, William Storm opened the debate concerning the preservation of audio documents proposing his two "legitimate directions" [23]. The first proposal aims to "the perpetuation of the sound of the original recordings as it was initially reproduced and heard by the people of the era" (Type I) [134]. In Storm's opinion, using the original equipment to rerecord the tape is possible to have a historically faithful copy of the sound, according to the historical conditions and technology of their era. Thus, the original recording and reproduction systems should be documented, too. The second proposal extends the first one, trying to achieve a more ambitious objective: to reproduce "the live sound of the original performer" (Type II) [134].

About twenty years later, UNESCO reports a new approach in [20] highlighting the motto "Save history, not rewrite it". This report was deeply influenced by the Schüller works [126]. In his paper, Schüller presents the technical and artistic analysis of the original carrier as the starting point for a rerecording. The proposed procedure limits the audio processing during the preservation phase in order to obtain the highest quality of the signal, abandoning the constraint to use the original equipment. An important classification is presented in [20] concerning the signal alterations grouped in *intentional* and *unintentional*. The first type of alterations includes equalization and noise reduction systems. The latter can be further splitted into two groups: (1)

imperfections derived from the original recording techniques and (2) the ones due to the mis-alignment of the recording equipment. Concerning the compensation of these alterations, three types of strategies were presented: the first one is similar to Storm's first approach and consists of a rerecording of what was originally heard. The second strategy proposes a preservation of the recording as it was produced: intentional alterations have to be compensated and the best modern equipment has to be used in order to minimize the replay distortions [20]. Any other kind of processing are not allowed. The last one allows to compensate alterations to reduce imperfections caused by recording techniques such as needle noise, rumble and tape hiss [126]. Although with some modifications, the second strategy is the base of the methodology developed at Centro di Sonologia Computazionale (CSC), the Sound and Music Computing laboratory of the Department of Information Engineering at University of Padova, where the author worked during his Ph.D. According to this methodology, the restoration is only allowed if it is necessary to optimize physical condition of the carrier before signal extraction. Only the intentional alterations (such as equalization, see Section 2.2.2) should be compensated. This position differs from what Schüller suggests. But unintentional alterations should not be compensated, since they witness the true history of the transmission of the audio document, representing an information related to the original context. In [23] and [67], the methodology is fully described which fulfill the fundamental points of preservation practice arose from the debate:

- "accurate, verifiable, and objective" procedures;

- measurements based on an ideally objective knowledge;

- modern playback equipment, fully compliant with the format specific parameters of the recordings;

- a careful documentation of all measures employed and of each manipulation applied (ensure reversibility).

For each analog audio document, a digital preservation copy is the result of the digitization process, which consists in a organized set of data and metadata that groups all the information represented by the source document, stored, and maintained as the preservation master. The main content is the audio signal, digitized and stored in high quality format, although without a sufficient set of metadata, all the digitization work can be compromised. By considering the long-term preservation, robust metadata helps to maintain the reliability, accuracy and authenticity of the digital document. One digital document without certain provenance can compromise the reliability of the entire archive, nullifying the digitization campaign with incalculable loss of time, money and even cultural materials (in case the originals became unaccessible) [67]. By thinking in a long-term perspective, it is highly probable that the original carriers will disappear or will not be readable. Moreover, it will not be possible to go back and retrieve missing or inadequate metadata as well as the original signal. For this reason, it is important to document also the digitization process as well as all the modifications of the digital data and metadata in order to reconstruct the history of the audio document.

However, a digitization that only focuses on the audio document (with its metadata) means a great loss of information, causing philological and musicological problems [24]. In order to

maintain the original context, several contextual information must be preserved in the preservation copy. For audio recordings, the content-independent materials are miscellaneous and of different nature, such as information written on the edition containers (envelopes, cases and boxes), carrier, or possible attachments (texts, images), physical conditions, intentional alterations and corruptions. In the preservation copy, contextual information can be stored using a combination of textual, photographic and video documentation. For example, photos are useful to report accurate information about labels, edition boxes and other attachments, as well as clearly visible carrier corruptions. Furthermore, information that cannot be directly represented in digital format has to be thoroughly documented in the descriptive sheet. An extreme example is the smell of a magnetic tape, which can indicate the presence of syndromes (mould, vinegar odor, etc.) [67]. Several research projects on metadata such as PREMIS[5] have been financed in the past years, but their focus was usually related to audio metadata and cataloging, without considering all these kinds of contextual information. Notwithstanding, this kind of information has a high impact on the accessing and the experiencing of historical audio documents.

Accessing and Experiencing Analog Audio Documents

Nowadays, the main debates and research efforts concern almost exclusively document preservation point of view, thus disregarding the essential aspect of music: the listening and the experience in its entirety. Some notable projects such as Europeana[6] exist but usually not consider contextual information as necessary, and are not adequate for musicological analysis. Moreover, international archives and digital libraries, usually provide Compact Disc player or "iTunes-like" audio players which are unfaithful to the peculiarities of the original (analog) audio document. This approach is highly influenced by contemporary pop music and by the modern way of life A musicological analysis with inadequate tools or incomplete information can lead to historical forgeries, difficult to detect. The main problems affecting the consumption of historical audio documents can be summarized as follow [36]:

1. impossibility to access the original analog audio document (for the motivations detailed in the previous section);

2. impossibility to access the original analog replay machine;

3. lack of a specific methodology to experience historical audio documents;

4. lack of adequate tools that consider the contextual information and could substitute audio players borrowed from other genres, artistic forms, or sectors;

Given the unavailability of the original audio document, the main reference is the preservation copy (the new master), whose bibliographic equivalent is the facsimile or the diplomatic copy, able to preserve the signal along with contextual information as well as the metadata. However, a complete high quality preservation copy can weigh several gigabytes, which can be excessive

---

[5]`loc.gov/standards/premis/` (Retrieved September 27, 2018)
[6]`europeana.eu/portal/en` (Retrieved September 28, 2018)

for several systems, in particular for web services. The relief is the creation of an *access copy*, which derives from the preservation copy, but has lower-quality audio files and/or a reduced amount of metadata and contextual information. There are several reasons to create a copy, for example, as adjunct to the catalogue, to help researchers decide what documents they wish to study. However, a copy of average/good quality may be also acceptable for access *in situ* [23], for the musicological study as long as the available set of data, metadata and contextual information is enough to maintain the integrity (from the philological point of view) of the audio document.

Unlike born-digital audio files, historical analogue audio documents are indissolubly linked to their physical carriers, on which they are recorded, and to the audio player (such as a gramophone or a tape recorder/player) which can strongly impact the listening experience [102]. In some case, the peculiarities of a carrier can heavily influence musical works. Therefore, they must be considered during the musicological analysis (see Section 2.1). Audio recorders/players, like audio documents, are characterized by a rapid change and improvement of the technology, and a fast obsolescence of their hardware. In addition, the knowledge and the skills of early audio technicians related to these machine are likely to be lost. The digitization of the original medium breaks the connection between the carrier and the audio signal. Nevertheless, scholars can only work on digitized copies of audio documents because usually the original carriers and the related playback devices are not available or even missing. Sometimes, the contextual information is not enough to reconstruct this bond. From a philological point of view, it is interesting to study the audio document in relation to its original recording/playback device. In this sense, it is important to preserve the behavior and peculiarities of such devices [36]. These features have to be provided in the audio player through which a scholar studies the audio document.

The preservation of electrophone devices is a new and interesting field, with many unanswered questions and open problems [67]. In particular, the devices should be preserved not only for museum exhibitions, but also to safeguard their functionality. The active preservation of the audio documents detailed in the previous section can be transposed in the field of recording/playback devices.

The active preservation of electrophone equipment can represent a major technological and scientific challenge, because it requires to analyze, understand, and simulate the behavior of complex devices, assembled from several components, some non-linear, some having partially unknown characteristics. In this sense, the virtual device should have the same role of an access copy in the field of audio recordings preservation. The aim is to build a human-computer interface able to reproduce a replay machine, offering the same functionality and allowing the same gestures of the original analogue device. Such interface should also enable a listening experience in agreement with the first "legitimate direction", namely, "the perpetuation of the sound of an original recording as it was initially reproduced and heard by the people of the era" [134].

Opportunities and open problems

The benefits of the digitization process are numerous and indisputable. First of all, the digitization can be used as a mean to save historical audio documents which, are endangered by several threats deriving from obsolescence, syndromes, degradation, etc. Digitization also enables the access to materials that were previously unavailable or only partially/locally available [23]. This

means to valorize such an important part of the cultural heritage, enabling and promoting the use of historical audio recordings. The creation of a complete set of metadata and data enables not only the use of a single audio document, but can help a "virtual re-unification" of collections preserved in different institutions, thus facilitating the study of a repertoire in its original context.

The digitization of significant amount of recordings also provides a big opportunity from the computer engineering perspective. The creation of solid corpora and datasets is essential for the implementation of data-driven research tasks in several fields, such as MIR and computational musicology. In this way, new kinds of computational studies and methodologies can support the traditional musicological research.

Nonetheless, as presented in the previous sections, during this process the document history may be distorted and the documentary unit could be broken, with the result that the digital copy is not reliable from the authenticity point of view. An appropriate methodology that considers the digitization of all data and metadata could not be enough. The process is not completely automatized and sometimes it is influenced by human subjective choices. Furthermore, several problems derive from human attention in repetitive tasks, which is generally required during the digitization process of big amount of recordings and by archival routines [117]. In this case, artificial intelligence can help operators, reducing its cognitive load and minimizing the introduction of unwanted errors into the system. The result is an increase in the reliability and accuracy of the process. Starting from the analysis of digital copies, an automatic algorithm concerning the feature extraction and classification can discover peculiarities related to the carrier and help deciding the necessary actions to be performed by the operators. These kinds of algorithms could also be the base for quality control systems applied to the digitization process [102].

The musicological analysis can also be affected by the same problems. Thus, such automatic tools are a good opportunity for the musicological studies, too. Furthermore, new tools can be developed in order to exploit the rich set of contextual information.

The issue concerning the experiencing of the digitized audio recordings also has to be tackled. Some commercial software package that reflect some characteristics of analog players (e.g., a gramophone) exist, but the virtualization can be incomplete and inaccurate. Since new interfaces have to be developed this represents a good opportunity to study and formalize the knowledge related to old analog machines, as requested in [55]. This research can include a study in HCI, that can potentially disrupt the way on which musicological analysis are performed. This is possible by only studying an *ad hoc* solution which incorporates all the requirements and features concerning the data, metadata and cultural context related to analog audio documents.

1.2.2 Historical Musical Instruments

Communicating and valorizing historical musical instruments can be particularly difficult, especially in the context of a museum exhibition. Presenting a musical instrument to the general public is a complex task, because of its multi-faceted nature. It is necessary to effectively communicate several aspects related to its cultural context, namely the history, iconography, acoustics, musicology, etc. Furthermore, in order to enhance the comprehension of a musical instrument and the visitor experience, listening its sounds and the music it can produce is essential. While respecting the original context, digital technologies can help in this task, making the museum

visit an active experience. In the last years, multimedia installations have been increasingly used in museum practices [110], altering the museum experience [133]. This way, a museum visit becomes both educational and entertaining [78]. Some researchers warned about the risk concerning the use of media technologies, which entertains at the expense of accuracy, distracting from real knowledge and undermining the educational experience [133]. This must be avoided: digital technologies have to valorize the artifact and not to replace it. The technology does not represent the starting point for the creation of a multimedia installation. The work has to start by analyzing the artifact and its context.

As mentioned above, in order to enhance the comprehension of a musical instrument and the visitor experience, it is useful to listen to the sounds and the music it can produce. Here are some examples of some museums that tackled this problem. The Cité de la Musique museum in Paris[7] valorizes the musical instruments included in its collection by using an audio guide. Through this device, the visitors can listen the sound of the musical instrument that they are watching. In this case, an approach of passive listening is used.

However, the visitor experience should be enriched through an active participation. It is desirable to have the opportunity of playing the musical instrument, such as in the case of the multimedia museum-house of Giuseppe Verdi[8] in Parma (Italy). Tablet and earphones are provided to the visitors of Verdi's house. By tapping in some defined points, the visitor can listen anecdotes, his music and play some related videos.

Museo Interactivo de la Mùsica in Malaga[9] (MIMMA), which mainly collects musical instruments, employs another approach. It gives the opportunity to touch and listen the sound of some instruments. Morover, it offers videos showing how to play the instruments as well as interactive installations that guide the visit through gamification.

A different approach is adopted by the European Music Archaeology Project (EMAP)[10] which reconstructs high-quality physical replicas of more than 60 ancient instruments, which are presented to the public that can try out, touch and play them in order to experience their sounds.

A last example is a research project of the Music Instrument Museum in Milan, where the original electronic instruments of the Studio di Fonologia Musicale (RAI, Milan) are exhibited [108]. During the 1950s and 1960s, this was one of the leading studios in Europe for the production of electro-acoustic music, together with Paris and Cologne, and where some devices were specially designed for a new type of music. In order to understand this kind of electrophone instrument, it is critical to listen the peculiar sounds that they produce. For this purpose, during the EU founded project DREAM[11], coordinated by CSC, an installation was realized re-creating the electronic devices, such as the oscillators and filters. The visitor can interact with the installation, which presents the same aspects and controls of the original instruments, by playing and producing electronic music [35].

An unique correct approach cannot be presented, as it depends from artifact to artifact. How-

---

[7]`philharmoniedeparis.fr/en` (Retrieved September 28, 2018)

[8]`casanataleverdi.it/en/` (Retrieved September 28, 2018)

[9]`musicaenaccion.com/mimma/` (Retrieved September 28, 2018)

[10]`emaproject.eu` (Retrieved September 28, 2018)

[11]`dream.dei.unipd.it/` (Retrieved September 28, 2018)

ever, to the author's knowledge, there is no prominent shared methodology to develop a multimedia installation. According to [96], a museum design culture is necessary to draw the role of the artifacts of museums and galleries with respect to the visitor experience. Design can be considered a mediator [148] between different disciplines (such as engineering, history and archaeology) and different actors (such as researchers, and the museum visitors). Due to this complexity, the visitor experience should be valorized under several aspects by taking into consideration three different design disciplines:

- *Interaction Design (ID)*, which role is facilitating interactions between humans through products and services [122];

- *Product Design (PD)*, which concerns the design of physical products by considering the object functions, aesthetics, ergonomics, usability and manufacturing possibilities;

- *Graphic Design (GD)*, which is the medium through which information is communicated.

This work tackles the lack of methodology in a multidisciplinary way, by considering the artifact and its cultural context as the starting point for the design of a multimedia installation.

1.3   Objectives

In the previous sections, several issues have been presented concerning musical heritage, and in particular analog audio recordings and historical musical instruments. This thesis tackles some of them by working on three main case studies.

In order to study the issues arising from the digitization, access and experiencing of analog audio recordings, as well as the opportunities provided by information engineering, the focus of the first case study is put on analog magnetic tapes. In particular, the case study considers the tape music genre, where the issues concerning the physical carrier are exacerbated. Composers of this genre worked directly on the tapes, physically cutting and pasting their parts, sometimes taking even notes on them. As explained in Section 2.1, this case study has relevant implications on the musicology, philology and archivists point of views, since in some cases composers did not provide the scores. Hence the tape results to be the artwork i.e., the final product of the creative process, which must be preserved to its fullest [114, 36]. The digitization process risks to brake the connection between the musical work and its carrier. A detailed study is proposed in Section 2.2 concerning the digital preservation copy of this kind of audio documents by highlighting the importance of the contextual information. A particular kind of contextual information is discussed and analyzed in Section 2.3 showing the opportunities deriving from applying artificial intelligence in their analysis. This preliminary work aims to open the way to a series of automatic tools to help musicologist and audio technicians in their work. To follow, Section 2.4 has the objective to validate the potential of machine learning techniques to recognize a specific aspect related to magnetic tapes: the equalization. Here again, the purpose is to relieve the audio technicians from arbitrary choices. Considering the peculiarities of tape music, the philological problem of stemmatics in digitized audio documents is tackled in Section 2.5. An innovative approach based on the phylogeny is used and evaluated with an experiment, with the purpose to determine the suitability for this task. Concerning the access and consumption of digitized audio recordings, once again, tape music provides an exceptional case studies. The objective is a methodological definition of the requirements for a faithful access to digitized tape music recordings to access and interact with digitized audio documents, which consider contextual information and its relation with the carrier as well as the replay device. This methodology is proposed in Section 2.6, in which some applications are proposed. In some web and mobile interfaces a virtualization of a tape recorder is proposed for a faithful experience of the audio document. The approach was validated by experts. Furthermore, a first prototype of web application is presented with the aim to concentrate relevant part of the knowledge acquired in this work in a single interface, studied on the base of a group of musicologists' requirements.

The second case study is presented in Section 2.7: a corpus of Arab-Andalusian music. A research corpus suitable for computational research opens new opportunities to musicological studies by applying data-driven analysis. According to [129], the creation of a corpus suitable for this kind of approach requires to consider five criteria. After the presentation of the peculiarities concerning this music tradition, the corpus is analyzed on the base of these criteria. Furthermore, several applications are presented with the aim to analyze the data and metadata of this corpus. Based on this application, a first approach of computational musicology applied to Arab-Andalusian music is proposed in Section 2.8.

The third case study concerns an historical musical instruments: an ancient Pan flute exhibited in the Museum of Archaeological Sciences and Art (MSA) at the University of Padova[12]. The final aim of the research project was the create a multimedia installation that would valorize this precious artifact and enable museum visitors to interact with it and its history. The case study provides the opportunity to study a methodology suitable for the valorization of this ancient musical instrument, but also extendible to other artifact or museum collections. In Section 3.2, both the methodology and the resulting multimedia installation are presented followed, by their assessment by a multidisciplinary group of experts.

---

[12]`beniculturali.unipd.it/www/servizi/museo/` (Retrieved September 28, 2018)

Chapter 2

Historical Audio Documents

2.1 Case study: tape music

A prominent example to highlight the importance of analog audio documents is *tape music*. Since the 1950s, this genre revolutionized the music creation process, becoming one of the most important cultural phenomena of the period [65, 115, 15]. Its peculiar working method was made popular by American composers such as John Cage and Steve Reich of the Columbia-Princeton Electronic Music Center, and European ones, such as Luciano Berio, Henri Pousseur, Bruno Maderna, and Luigi Nono working at the Studio di Fonologia Musicale of RAI in Milan [115]. Later, tape music evolved along with technologies for music post-production, embracing most genres and aesthetics trends of recorded sound arts.

This kind of music cannot be set in conventional notation: the musical text is non-existent, incomplete, insufficiently precise and transmitted in a non-traditional format. Tape music consists of the manipulation of fragments and samples of pre-recorded sounds in order to obtain a modern composition. The manipulation physically involves the tape: the samples are pieces of tape which are cut, altered, edited, superimposed, etc. [60] in order to create new sounds. In this way, the composer also became the luthier and the performer of the completed product recorded on magnetic tape, which can be considered as a *unicum*. This *unicum* is the result of an handmade work. For this reason, tape music ceases to be only an allographic art [75], performed through the contribution of many actors (i.e. the composer and the performer) and acquires characteristics of autographic art [75], where a product complete in itself is available at the end of the creation process (similarly to other art forms such as painting, sculpture, etc.) [36]. In fact, the uniqueness of tape music works tackles a well-known problem in the visual arts field, such as the attribution and the generation of different versions (called witnesses, in the philology field) [144].

Tape music is the most challenging musical content to preserve and analyze on a magnetic tape, as it presents issues that do not exist in other scenarios [67]:

1. tapes were physically manipulated with cuts, splices, etc.;

2. annotations crucial for the music performance were sometimes applied on the tape itself, and a formal score was seldom produced;

3. as a general rule, the composer considered the technical possibilities and the constraints of the specific recording/playback device, exploiting them, sometimes overcoming them by modifying the device with the help of technicians;

4. the presence of concrete and/or electronic sounds together with acoustical instruments makes it difficult to distinguish between audio corruptions and intentional alterations.

For these reasons, tape music represents a paradigmatic case of recorded sound art and its peculiarities provide an important food for thought in the musicological, philological and preservation research fields [33, 11, 36]. Furthermore, magnetic tapes are the most demanding audio media to handle, because of the unrivaled possibilities for manipulation they offer, and because of their frailty and short lifespan. As this work is perfected for the most challenging musical content on the most demanding audio medium, it can be extended to other media and genres, e.g, Western classical music, opera, ethnomusic, speech corpora.

2.1.1   Open works

From the musicological and philological points of view, several interesting artworks are realized in an open form. In classical musical composition, the performer reproduces the idea of the composer that is well-defined into a score. On the contrary, some music works provide a considerable autonomy to the performer which can choose how to play the work. In this case, the latter one is not merely free to interpret the composer's instructions following the score such as in traditional music, but he must impose his judgment on the form of the piece (e.g deciding how to group the sounds).

An example is *Klavierstuck XI* where the composer, Karlheinz Stockhausen, gives a single large sheet of music paper with a series of note groupings to the performer which has to choose the order of these groupings, changing the combinative structure of the piece. Another example is the *Third Sonata for Piano* by Pierre Boulez. The first section (*Antiphonie, Formant 1*) of this work is made up of ten different parts corresponding to ten sheets of music paper. As a stack of filing cards, these ones can be arranged in different sequences providing the possibility to create several versions (though some possible permutations are not allowed). A particularly representative example of musical open work is *Scambi*, an analog tape work created in 1957 by the Belgian composer Henri Pousseur at the Studio di Fonologia Musicale della RAI, realized by using a special equipment realized by Alfredo Lietti [35]. Using a specific process, called dynamic filtering, the composer was able to extract temporal structures from noise and to further process them with different parameters. As result, Pousseur produced 32 sequences recorded on tapes that can be arranged – according to a certain order and overlapping rules. All these rules were defined thinking to use of the original analog replay device: the tape recorder. Thus the original replay device became part of the artistic works itself creating a philological problem for musicologists and archivists which, nowadays, respectively analyze and manage the digital versions of these artworks. The original artworks are usually not accessible and, furthermore, the original replay device is not available. These kinds of problems are correctly approached

by the Scambi Project[1], that studies the use of open forms in electro-acoustic music, especially in compositions produced between 1950 and 1980 [36]. A new approach to tackle this kind of problems will be proposed in Section 2.6.

---

[1]`scambi.mdx.ac.uk/` (Retrieved September 16, 2018)

2.2   Digital preservation copy

In the last 20 years, one of the main research fields of CSC was being the preservation of historical audio documents [32, 151]. The multi-faceted problem of audio preservation requires a multidisciplinary approach, necessary to preserve and exploit the potential of this huge part of the cultural heritage. In these years, a methodological framework has been perfected, tailored on magnetic tapes needs [23, 67] and tested during several international preservation projects in collaboration with some of the most important international and national archives, such as Paul Sacher Stiftung (CH), Fondazione Arena di Verona (IT), Historical Archive of the Teatro Regio of Parma (IT), Luigi Nono Archive (IT), etc.

As explained in Section 1.2.1, with active preservation, the audio document loses its materiality and its content is preserved in a digital form. Nevertheless, many aspects have to be considered during the digitization of a tape [117]:

1. the object' material structure: the set of its physical-chemical components, the technology, the production system (acoustic, electro-acoustic, magnetic), the audio and playback format (such as speed and equalization);

2. the primary information related to the audio information contained in the recording;

3. the secondary (or ancillary) information [24, 23]: notes on the box, noise signals characterizing the recording system, discontinuities on the carrier (corruptions, splices, signs, etc.);

4. metadata;

5. the history of the document transmission (storage, duplication, etc.).

All these information, data and metadata have to be stored in the digital preservation copies, which is "the artifact" designated to be stored and maintained as a preservation master for the ages, when the original copy will disappear or will not accessible anymore. The bibliographic equivalent of this copy is the diplomatic copy or the facsimile [144]. Each preservation copy maintains:

- the audio signal of the document digitized and stored in high quality (sampling at 94 kHz and resolution at 24 bit) not compressed stereo or mono files (depends from the number of tracks of the tape, although it is preferable to always use the latter one);

- the photos or scansions of the carriers, the boxes, notes and every other documents related to the audio document;

- the video of the tape that flows on the heads of the tape recording during the digitization which role will be analyzed in-depth in the next section;

- all the metadata related to the document, including also the description of the initial status of carrier, potential physical restorations of the carrier and any other information related the digitization process.

Thanks to a continue monitoring of the audio content, the audio technicians can detect potential changes of speed in the tape. In this case, the tape is digitized a second time in the same preservation copy with their relative metadata and a new copy of the video. Several other alterations can affect the magnetic carrier (and so the digitization) and they need to be handled correctly and carefully described during the digitization process. In the next section, the main alterations concerning the magnetic carrier are described. During the digitization process, the replay device need also to be correctly configured. The main parameters are equalization and speed. In Section 2.2.2, the "equalization problem" is presented and discussed.

2.2.1 Tape discontinuities

In literature, several works enumerate the degradation factors of a tape [79, 21] Before the digitization, some of them are just recognizable with a first visual inspection of the operator such as for the following alterations: blocking, leafing, windowing, spoking, and embossing. These kinds of alterations can be communicated in the preservation copy with photos and with a description and then restored. Unfortunately, some degradations and syndromes are not detectable during this phase, but only during the tape replay (therefore during the digitization) endangering the audio signal transferring. Furthermore, the effects are localized in a small areas of the tape. For this reason, the replay monitoring is essential to detect them. Once detected, an alteration can be restored and the digitization redone in order to hand down a correct digital copy of the tape.

The main alterations detectable during the replay of the tape are [21, 117]:

Cupping an abnormal flexure of the tape surface across or along its width, due to different rates of shrinkage along the substrate and recording layers.

Damages on tape edges occurring when the edges do not appear flat or straight.

Riffles formally known in [105] as *Kink* if it consists in a crease on a layer of the tape or *Wrinkle* if consists in multiple creases in the tape.

Tape contamination and dirt presence of mold, powder, crystals, other biological contaminations or other particularities.

Interlayer adhesion stickiness of the surface of one layer to the back of the succeeding layer which could be cause of wow and flutter.

Gummy deposits presence on the tape of gluey substances which gather on the heads and guides of the playback machine during the tape replay.

Backcoat and magnetic shedding the first one involves backcoat particles coming away from the substrate and accumulating on surfaces in contact with the back of the tape. Loose debris can impair playback quality, depositing on the playing surface of the adjacent layer. The second phenomenon, due to a loss of cohesion, entails magnetic coating particles coming away from the tape substrate and depositing on the heads and guides of the playback machine.

Brittleness frequently with cupping, entails tape breaks easily.

The alterations listed above can be considered unintentional. Other kinds of alterations derive from the human intervention. The main one is the *splice*. According to [105], a splice can be defined as a small piece of special adhesive tape used to join two pieces of tape (magnetic or non-magnetic) to form a single piece. A non-magnetic tape is flexible plastic or paper strip that can be spliced to either end of a roll of recording material [3]. According to [64], an ideal splice does not cause an audible disturbance during the playback. However, in real situations, it is easy to find old splices that introduce perceptible disturbances. The primary cause of disturbance is represented by cut angle. The best angle is 45° and, raising the angle degrees, the risks of electrical disturbance increase. If the angle is lower than 45°, the tape becomes vulnerable to breaks. Another important kind of intentional alterations are signs or words which are written on the back of the tape (magnetic or leader) or on the adhesive strip of the splices (generally referred as *mark*). They are critical from the musicological point of view, since they could denote instructions for the performance of the piece or indication of particular sound events. Henceforth, the term *discontinuity* will be used to indicate all those alterations of the carrier from its original manufacture state detectable during the flowing of the tape [117]. Sometimes manufacturers printed their brand name or logo on the back of the tape itself. *Tape brands* will be considered as discontinuities, even if are not alterations.

The use of splices and signs become crucial in the tape music, where splices and signs are part of the creative process. Different samples of synthesized and/or elaborated sounds, recordings or part of them were joined together in order to obtain new sounds and effects. For example, leader tapes were used to add a pause or to signal new events or units. Therefore, the study of the magnetic tape assembly operations, and so of the alterations, gets relevance in order to analyze the genesis of the work of art. At the same time, the necessity to discriminate intentional alterations from unintentional ones become indispensable. Unfortunately this assessment can be pursued only during the replay, monitoring hours and hours of tape. Despite the information reported on containers, labels, attachments as well as the first kind of alterations of the physical carrier described above, which can be transmitted and stored as static images, tape discontinuities are strictly connected to the audio signal and required to be digitized differently. The solution proposed in [23] is a video of the tape during the replay shooting the back of the tape flowing on the head of the tape recorder. In order to be synchronized with the high quality audio, this video also needs the recording of the audio, even if in low quality. In addition to the contextual information concerning the tape discontinuities and the possibility to discriminate intentional from unintentional alterations, the video recording also offers the possibility to verify other irregularities in the playback speed of the tape which cause changes in frequency, such as wow or flutter. This choice implies that, if the tape is ruined only the front side, alterations cannot be detected and reported. These kinds of videos represent the input data of the software tool presented in Section 2.3.2, which is able to automatically locate and classify the discontinuities.

2.2.2 Equalization

The term "equalization" can be used to indicate any procedure that involves altering or adjusting of the overall frequency spectrum response of the audio signal. Equalization is used for several purposes. For example, during the audio mastering, equalization is used to improve the aesthetic of sounds.

The concept of filtering audio frequencies dates back at least to the 1870s. It was first applied in harmonic telegraphs, and then later adopted in analog audio recordings [142]. The equalization role was paramount since it allowed to compensate physical limits derived from analog technologies and their carriers. Henceforth, unless otherwise specified, only this kind of compensation will be considered as equalization.

Its working principle is the following. A pre-emphasis curve is applied to the signal which is written in the analog carrier, and an inverse post-emphasis curve is applied during the reading phase. Thus, the resulting output signal maintains nearly the flat frequency response of the original input [98], but at the same time, it contains an extension of the dynamic range [68] and an improvement of the SNR ratio [30]. This technique is adopted from several analog audio technologies for two main reasons: "the limited dynamic range of audio systems and the fact that music sources produce more energy in the low-frequency region where the ear is less sensitive to noise" [68].

Historically, the adoption of these techniques was not uniform, and several different standards were adopted by record manufacturers. For example, for grooved media, the situation was chaotic until the introduction of standards promoted by Recording Industries Association of America (RIAA) in 1955 [51]. Before this date, dozens of standards existed. In order to faithfully reproduce (and so to digitize) recordings of these preceding years, it is necessary to tackle what is referred to as [51] the "equalization problem".

When analyzing magnetic tape technology, the same problem arises. Several standards exist, and during playback and digitization, this must be considered in order to obtain a faithful listening experience and a trustworthy preservation copy. A complete list of the most common equalization standards for audio replay of analog tape can be found in [21]. Equalization standards are usually referred to with the acronyms of the organization that proposed the standard itself. Historically, different standards were most widespread in Europe and in America. The most prevalent European standard was IEC1 from the International Electrotechnical Commission, alternatively called *CCIR* by the acronym of the *Comité Consultatif International pour la Radio*. In America the most prevalent standard was IEC2, also referred *NAB* from the American National Association of Broadcaster. Henceforth these equalization standards will be referred as CCIR and NAB.

The equalization standards are strictly connected to another parameter that must be correctly configured before the equalization setting: the replay speed. There are six standard speeds: 30 ips (76.2 cm/s), 15 ips (38.1 cm/s), 7.5 ips (19.05 cm/s), 3.75 ips (4.76 cm/s), 1.875 (4.76 cm/s) and 0.9375 ips (2.38 cm/s), but a tape recorder with all these speeds in the same machine does not exist. The most common are 15 ips and 7.5 ips [21]. For these reasons, CCIR and NAB equalizations will be considered at these replay speeds in the following analysis.

A post-emphasis curve could be expressed as a combination of two curves described with the

following formula:

$$N(dB) = 10\log(1 + \frac{1}{4\pi^2 f^2 t_1^2}) - 10\log(1 + 4\pi^2 f^2 t_2^2) \qquad (2.1)$$

where f is the frequency in Hz and t_1 and t_2 are the time constants in microseconds [81]. An alternative mathematical representation of the formula is:

$$N(dB) = 20\log_{10} \omega t_1 \sqrt{\frac{1 + (\omega t_2)^2}{1 + (\omega t_1)^2}} \qquad (2.2)$$

where $w = 2\pi f$ and f is the frequency [106]. The two time constants in milliseconds describe the equalization curve, but in some case t_1 is ∞.

Equalization	Speed			
	$t_1(\mu s)$	$t_2(\mu s)$	$t_1(\mu s)$	$t_2(\mu s)$
CCIR	∞	70	∞	35
NAB	3180	50	3180	50

Table 2.1: *The time constants of the CCIR and NAB post-emphasis curves at 7.5 ips and 15 ips.*

For CCIR and NAB at 7.5 and 15 ips audio tape recordings, the time constants are reported in Table 2.1 and the characteristics of the related equalization will be analyzed in this paper. Figures 2.1(a) and 2.1(b) present the frequency responses of pre- and post-emphasis curves respectively for NAB and CCIR equalization at 7.5 ips. As can be seen in Figures 2.1(c) and 2.1(d), an incorrect juxtaposition of the pre- and post-emphasis can significantly alter the spectrum, compromising the faithfulness of digitized audio documents.
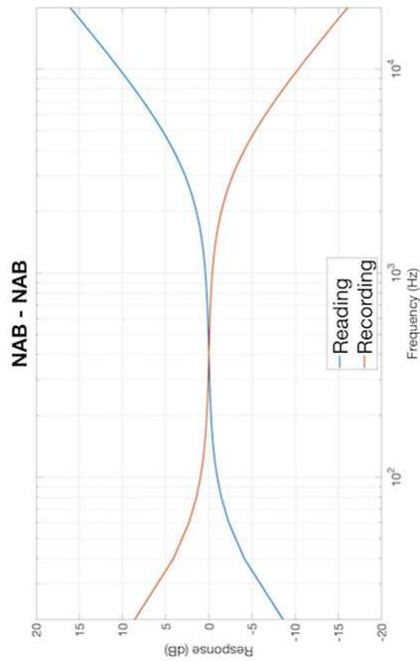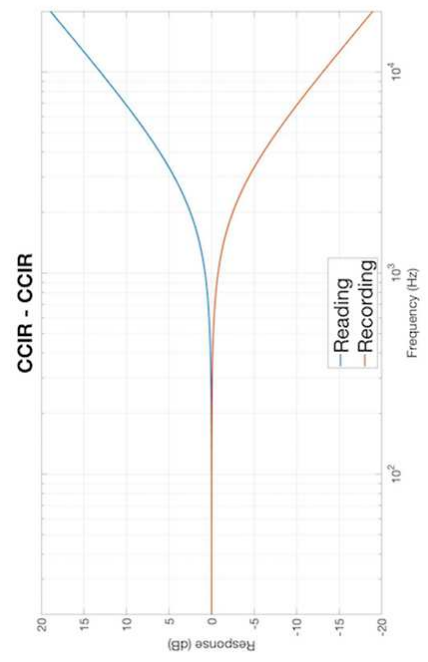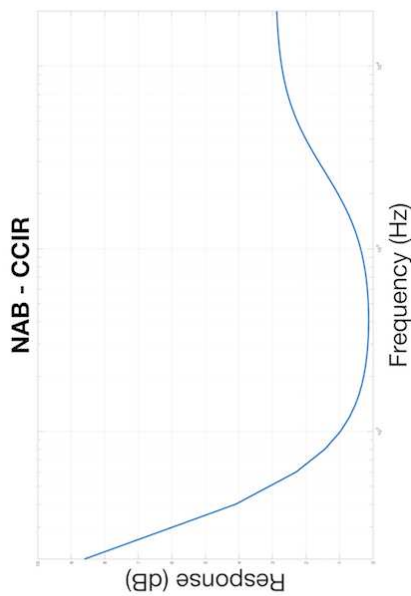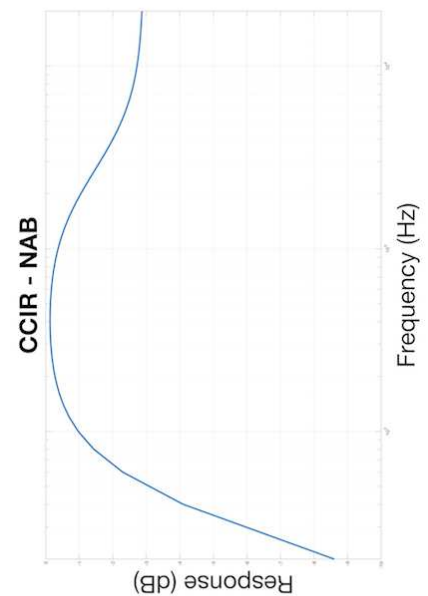
Figure 2.1: *The frequency responses of (a) NAB and (b) CCIR pre- and post-emphasis curves and their incorrect juxtapositions: (c) NAB - CCIR and (d) CCIR - NAB.*

2.3   Computational discontinuity detection

Both the creation and the study of a digital preservation copy share a common problem: the human attention [117]. Discontinuities can be detected only by carefully monitoring the entire tape, during the digitization process or the musicological analysis. Because of so many hours of listening and the observation of kilometers of flowing tape, the work of a technician or a musicologist can become chaotic. Some of the disorder patterns are due to those present (and well-known in literature) in the transcripts made by the *servus a manu scribes*, while others are not studied yet. The causes of inattention can be so many that the probability of an incorrect choice is high. This can lead to the creation of poor value preservation copies or incorrect analysis.

An automatic approach can help to avoid human errors of both audio technicians and musicologists. The idea proposed in the following sections is to analyze audio and video of the tape in order to extract information from the digitized copy, relieving audio technicians and musicologists from repetitive, tiresome, or otherwise error-inducing tasks. The creation of a summary report with all the discontinuities can be provided to professionals who can take decisions based on their high-level interpretation. Thus, these kinds of analysis can be considered a pre-processing of the information.

The next sections aim to investigate on possible sources of information, evaluate their feasibility and test algorithms able to find discontinuities on audio or video.

2.3.1   Splice recognition from audio

Alteration of audio signals and audio restoration are well-known topics in the literature [74, 34, 31] Some alterations of the signal of a digitized audio document are related to discontinuities of the carrier. In this section, it is presented an attempt to find a relevant feature able to discern one of the most frequent discontinuity: the splice. In case of a splice that joins a magnetic tape with a leader one, the presence of the splice can be observed and deducted from the spectrogram [144]. Where two magnetic tapes join with a homogeneous audio signal, the situation is different. As presented in Section 2.2.1, the cut angle of a splice is essential both to avoid electronic disturbance and for physical resistance. Using a cut at 45°, a fade-in/fade-out effects between the two extremities of the tape is created, whereas in the 90°splices, the risk of a mismatch between the two ones is high [64].

In order to characterize the audio cues of a splice, a first set of 40 splices has been created, in collaboration with Valentina Burini and Alessandro Russo. Half of them have a cut at 45°, whereas the other ones at 90°. Both sets are equally subdivided in splices created on virgin tape (VI) and other ones created in a tape recorded with a "silence track" (SI). In order to experiment the worst case, the tapes containing the splices were recorded and digitized at the fastest speed, 30 ips, in a professional recorder Studer A810 [135]. All the samples were digitized with a sampling at 96 kHz and a resolution of 24 bit in order to allow an analysis of frequencies over the hearing range.

The result of the test is unequivocal: all the 90°splices are clearly recognizable in the spectrogram whereas the 45°splices are not, with only one exception that is however weakly distinguishable. Magnetic tape recordings are generally free of clicks [73], but as can be seen in

(a)

(b)
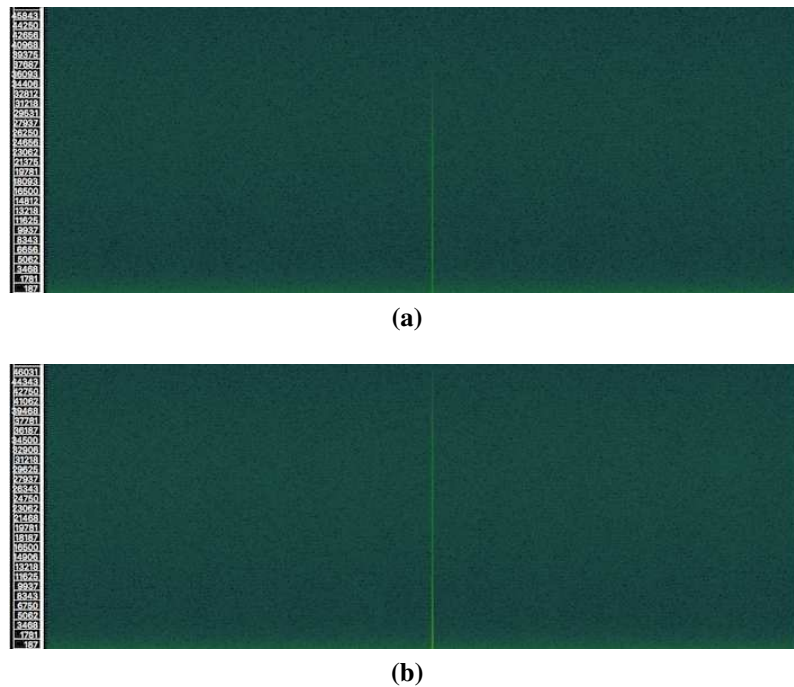
Figure 2.2: *The spectrograms of the two channels of the sample 7: a splice on virgin magnetic tape.*

Figure 2.2, for 90° splices a spike is visible in at least one of the channels of the digitized samples. The peaks are not uniform in all the samples but involve all the frequencies from zero to a variable maximum value, as can be seen in Table 2.2. Furthermore, some peaks greatly exceed the hearing range nearly reaching the maximum frequency range of 48 kHz.

Other 80 samples were used as second test. For this test, a tape has been recorded with four categories of sounds: instrumental acoustic music, speech, music made with electric instrument, electronic music. A splice between two strips containing two completely different recordings (join together to achieve a pursued effect or musical intent) is relatively easy to find considering the non-homogeneous signals In case of two homogeneous strips spliced together (as could happen in case of broken tapes), the work becomes harder. In order to have the worst case, the recorded tape has been cut, and its extremities reattached again in the same position.

Table 2.3 presents the results of the two tests. While for the 45°samples, the results confirm the previous test (as expected), for the 90° ones, the number of recognized peaks drastically decreases: the number of discernible peaks moves from 100% to the 42.5%. In the first step, only the 60% of the peaks exceed the 20 kHz. So, a plausible explanation of this behavior could be the fact that these peaks have lower power with respect to the surrounding frequencies texture. In conclusion, these preliminary tests prove that these peaks are not a serviceable feature to recognize splices with cut at 45°. On the other hand, the 62% of 90°splices are distinguishable. This overall result is better, but not adequate for a reliable recognition of a splice. For this reason, this attempt has been abandoned. A problem in the recognition of this kind of splices

Table 2.2: *The maximum peaks values of 90 degrees splices in the first test.*

VI 90° Samples	Channel 1 (kHz)	Channel 2 (kHz)	SI 90° Samples	Channel 1 (kHz)	Channel 2 (kHz)
1	7	0	1	17	19
2	0	47	2	15	22
3	43	44	3	22	20
4	0	19	4	16	22
5	10	24	5	16	17
6	9	10	6	14	20
7	34	44	7	13	12
8	6	34	8	13	22
9	13	23	9	13	23
10	3	19	10	26	19

Table 2.3: *Results of the two tests on splices recognition. Legend of the type of samples (Types): virgin tape (VI), silence (SI) instrumental acoustic music (AM), music made with an electric instrument (EI), electronic music (EM) and speech (SP).*

Types	45° Splices		90° Splices	
	Recognized	Not Recognized	Recognized	Not Recognized
VI	0	10	10	0
SI	2	8	10	0
AM	7	3	9	1
EI	0	10	3	7
EM	0	10	3	7
SP	0	10	2	8

also emerged in the experiment described in 2.5. An alternative approach using the video as input has been studied and is presented in the next section.

2.3.2   Discontinuity recognition and classification from video

As introduced in Section 2.2.1, the video is the best way to maintain contextual information related to the carrier. Discontinuities can be documented and synchronized with the audio signal, facilitating the musicological study. As presented in the introduction of Section 2.3, human attention related problem can affect the quality of this analysis: some useful information could be lost, or some physical degradation misunderstood. In order to facilitate the work of musicologist and audio technicians, an automatic analysis software is necessary. A preliminary work which aims to the creation of a dataset suitable for data-driven analysis of the discontinuities is presented in the next sections. This work, made in collaboration with Carlo Fantozzi and the supervision of

Sergio Canazza concerns three aspects: the extraction of discontinuity frames from the videos, the definition of discontinuity classes and the creation of the dataset.

Extraction of the relevant frames

In order to extract the discontinuities, each video is analyzed frame by frame by an original software written in C++ and based on the library OpenCV[2]. The videos comes from several years of preservation projects, and in order to have enough material, an old format of the camera has been maintained with a resolution of 720x576 pixels. This old format introduces a problem related to the interlacing. In order to resolve this problem, several techniques of de-interlacing have been studied such as blend, resize, weaving, temporal averaging blend and edge line average, with not sufficient results. Software such as VLC use the techniques named *bob* where odd (or even) half-frames are duplicated in order to obtain a stable image. In this case, duplicated information is not useful for the analysis. The mediocre result obtained required a drastically choice: to consider only odd lines and discard even ones.

The starting idea was to individuate relevant changes between two frames by comparing them with the OpenCV tool BackGroundSubtractorKNN. However this technique was early abandoned because of the high number of false positive (frame not relevant for the analysis). The common characteristics of this false positive were:

- maintaining the same position;

- not flowing in horizontal direction;

- usually few points compared to the normal dimension of a discontinuity.

A second version of the algorithm has these improvements:

- only shape with a height between the 1/12 and the 1/3 of the number of rows in the frame;

- only flowing shapes and so, concerning only the part related to the tape.

In order to obtain the last constraint, one imaginary line has been sketched at 2/3 of the frame, and only the shape that passes for this line has been considered. The result of this improvements is the complete removal of false positive, but at the same time, some small discontinuities were missing. Furthermore, the computation time for each video was too long.

The second version of the algorithm has a substantial improvement. The dimension of the tape has been estimated based on the first seven discontinuities. In order to reduce the computational load, the BackGroundSubtractorKNN has abandoned and substituted with a subtraction operator between the three RGB values of each pixel. In some cases, the bound of the tape was not correctly identify. The possibility to create a manual bound resolved the problem and further reduced the computation time. On average, the analysis of a 20 minutes video decreased from 23 minutes (with automatic boundary detection) to 6 minutes.

---

[2]`opencv.org/` (Retrieved September 16, 2018)

Classes

Based on the discontinuity classification of Section 2.2.1, nine classes were identified:

1. Leader Tape / Magnetic Tape (L-M) Splices (Figure A.4);

2. Magnetic Tape / Magnetic Tape (M-M) Splices (Figure A.5);

3. Brands (Figures A.6 and A.7);

4. End of Tape (Figure A.1);

5. Ripples (Figure A.3);

6. Damaged Tape (Figure A.8);

7. Marks (Figure A.10);

8. Dirt (Figure A.9);

9. Shadows (Figure A.2).

The first two categories distinguish the splices in two kinds: splices that join a magnetic tape with a leader tape and others that join two slices of magnetic tape. Brands may be the full name, or just a logo of the tape manufacturers and can change in size, shape, and color, complicating the classification task. If present, usually they are repeated in the whole length of the tape. The discontinuity End of Tape concerns the moment when the tape ends. In that moment, the tape unties the reel and loses tension. The most evident event is the detachment of the capstan [30] as in (Figure A.1). Ripples class groups all the alterations of the tape's shape, such as cuppings and damages on tape edges, as described in Section 2.2.1. The Damaged Tape class groups all kinds of damages on the tape' surface that are not related to its shape such as contamination, creases, etc. Marks can significantly differ from each another: this class groups signs, words or symbols written on magnetic, leader or adhesive tapes. Frames that show irregularities not related to physical damages, ripples or marks are grouped in the Dirt class. These are usually due to a dirty capstan, which smears the tape during playback, or to residues of adhesive tape. The class Shadows is not described in the Discontinuity Section as it is not related to the tape condition, but to external shadows or reflexes due to lights, objects or persons that have been moving near the tape recorder during the digitization phase. Actually, this class is not relevant from the musicological point of view, but it is inserted in order to discriminate moving shapes with nothing interest from frames belonging to Damaged Tape class.

The dataset

In order to use neural networks or other machine learning techniques, a large training set covering all the classes has to be created. Two training sets have been created: one for 7.5 ips and another for 15 ips videos. These videos come from several digitization projects of CSC[3].

---

[3]for copyright restrictions cannot be released to the public

(a)

(b)

Figure 2.3: *Two frames with a completely visible splice (a) and a barely visible one (b), respectively.*

In a pre-processing step, tens of thousands of images were extracted and around 40,000 potentially significant images have been manually labeled (or discarded in case of false positive). The majority of the detected discontinuities were Brands. This is due to the fact that if a tape is characterized by the brand, this will be repeated continuously with a regular rate (usually very frequently). In order to speed up this phase, two neural networks (7.5 ips and 15 ips) were created using fine-tuning modified versions of GoogLeNet, proposed in [136]. Initially, the training set was composed by around 30,000 brands plus other different kind of discontinuities, and the validation set by more than 3,000 brands. After a first observation, the number of brands has been reduced to 5,000 in order to make smaller the imbalance between the classes. Furthermore, the resulting classified frames were manually checked if the confidence was below 90%. Some class was further checked, such as the L-M Splices. In some case the splice was barely visible, as in Figure 2.3(a). Only completely visible splices (as in Figure 2.3(b)) were maintained in the dataset. The discarded ones were however saved because they could be useful for testing extreme situations. Another problem concerned the presence of multiple discontinuities in the same frame. The following criteria were created in order to resolve ambiguous situations:

Class	Elements (7.5 ips)	Elements (15 ips)
L-M splices	1,038	1,457
M-M splices	210	555
End of Tape	2,484	85
Brands on Tape	5,834	5,709
Damaged Tape	444	175
Ripples	6	0
Shadows	1,333	51
Marks	118	346
Dirt	8	9
TOTAL	11,475	8,387

Table 2.4: *The number of elements for each class of the discontinuity datasets.*

- If the frame contained a horizontal stripe of adhesive tape, it was assigned to the Marks class. The stripe was applied horizontally on the tape not to fix a damage, but to indicate a significant moment for the composer.

- If the frame contained any writings, it was placed in the Marks class.

- Frames where the capstan was moving away from tape because the playback is ending, were classified as End of Tape only when the capstan was fully released.

- A splice with writings was assigned to the Marks class.

- A splice with surrounding damaged tape was placed in the Splices class.

The final datasets totally have of 19,682 images, distributed as in Table 2.4. The use of real audio documents entailed some difficulties related to the balance of the training set. Discontinuities such as Dirt or Ripples were rarely found. For this reason the dataset is not equally distributed across the nine classes. However, the main classes are suitable for preliminary experiments applying machine learning techniques such as clustering and classification. At the moment, this dataset is being enlarged with the purpose to become suitable for neural network algorithms such as the one proposed in the extraction phase.

2.4 Equalization recognition

Often, speed and equalization standards are not indicated in the cover or in the notes of the audio document and this fact endanger the correct transfer of the documents in the digital domain. As stated in [51, 21], without reliable documentation or frequency tones, the operators involved in the digitization process have to choose the equalization aurally. However the differences between the equalization curves are subtle and do not involve all the audible range of frequencies, so during the digitization process it could be difficult, for the audio technician, to determine the correct one. The experiment described in [28] studied the capability of skilled and unskilled testers to discriminate digitized audio samples with different equalizations through a Multi Stimulus test with Hidden Reference and Anchors (MUSHRA) [83]. In each test correct and incorrect equalized samples were compared with a correct reference, and despite the reference several error occurred. Considering that the audio technicians who digitize a document do not have the reference, it is easy to understand how this task could highly error-prone, in case of a lack of documentation or indications from a musicologist.

To avoid subjectivity and therefore errors that can damage the integrity of the preservation copy, it is necessary to relieve this task to the technicians. The solution could be the use of automatic tools able to analyze the digitized audio signal, in order to find the original equalization with which the tape was recorded. The use of this kind of tools could be useful not only for audio technicians, but also for musicologists that can prove the correctness of the digital preservation copy of unknown provenance. The idea of the following experiments is the use of background noise of the tape in order to discriminate the equalization applied during the recording of the tape and that one used during the digitization process. With the supervision of Sergio Canazza and the collaboration of Edoardo Micheloni, several machine learning algorithms have been tested in order to determine and study their efficacy and reliability in the equalization recognition task.

2.4.1 Preliminary experiments

In order to evaluate the feasibility of machine learning algorithms in the equalization recognition, a preliminary experiment was created. The idea is to use the intrinsic noise deriving from analog recording and reading [54] to recognize the equalizations used in these two steps. In order to evaluate the validity of the noise as element to distinguish an equalization, two virgin tapes were recorded with white noise and a "silence track". The latter one was used to record only the noise derived from recording and reading. Both the two types of noise were recorded in 7.5 and 15 ips, obtaining the four datasets presented in table 2.5. Each dataset has four kinds of samples made alternating CCIR and NAB equalization in pre- and post-emphasis. The four resulting

Table 2.5: *The characteristics of the four datasets for equalization recognition.*

Recording/Speed	7.5 ips	15 ips
Silence	dataset A	dataset C
White noise	dataset B	dataset D

pairs are CCIR-CCIR (CC), NAB-NAB (NN), CCIR-NAB (CN) and NAB-CCIR (NC). For each pair, there are 300 samples of one second, and therefore a total of 1200 samples for each dataset. With the Matlab tool Mirtoolbox (Music Information Retrieval Toolbox [89]), 13 Mel-Frequency Cepstral Coefficients (MFCCs) were extracted from each sample. These features, originally developed for speech-recognition systems, performed well for a variety of audio classifications [14]. Moreover, they allow a low computational cost and a fast training. Each sample is therefore a vector of 13 coefficients.

The experiment tests two kind of machine learning techniques: cluster analysis and classification. For cluster analysis, two of the most used techniques of unsupervised learning were adopted: hierarchical clustering and K-means clustering. Several distance measures (i.e. Euclidean, chebychev, cosine, etc.) and linkage methods (i.e. average, single, etc.) have been tested (with the constraint of maximum four clusters) for the first technique while, for the latter, distance measures and number of clusters (from 2 to 4) were used. Considering the combination between the parameters, 232 tests have been executed: 188 tests (47 x 4) for the first method and 48 (12 x 4) for the second one.

Regarding the classification, the experiment tested three of the most common supervised learning techniques: Decision Tree, K-Nearest Neighbors, and Support Vector Machine (SVM). As well as for the unsupervised learning, tests based on these techniques differ for the chosen presets and parameters. Decision Tree tests use three kind of preset, concerning the maximum number of splits: Simple Tree (maximum 4 splits), Medium Tree (maximum 20 splits) and Complex Tree (maximum 100 splits). The six tested variants of K-Nearest Neighbors, which differ for the number of neighbors and distance metrics, are: Fine, Medium, Coarse, Cosine, Cubic and Weighted. The last method, SVM, was tested in five variants, which differ for the kernel function: Linear, Quadratic, Cubic, Fine and Gaussian. Each dataset was divided in a training set with 75% of the samples and a test set with the other 25%.

Clustering results

The best results of the several tests based on clustering analysis are the following:

- datasets B and D (white noise) were divided in three clusters, one with the samples with the right juxtaposition of the post-emphasis (CC and NN), and two separate clusters for respectively samples NC and CN;

- datasets A and C (silence) were grouped in only two clusters, one with all the samples with NAB post-emphasis (NN and CN) and one with the CCIR ones (CC and NC).

Most of the different combinations of distances and linkage methods of Hierarchical clustering are able to discern white noise samples, as can be seen in Table 2.6 where a clustering with Euclidean distance and centroid as linkage method is presented. In general, K-means technique does not work with these datasets. The only exception is the method that uses cityblock distance, where the samples are grouped in three clusters as for the other unsupervised learning method. Vice versa, for silence samples, K-Means algorithms work better using most of the distances,

Table 2.6: *The four clusters of white noise samples resulting from Hierarchical clustering algorithm with Euclidean distance and centroid as linkage methods.*

Cluster	Cluster 1				Cluster 2			
Distance	CC	CN	NC	NN	CC	CN	NC	NN
# samples	0	0	2	0	0	0	298	0
Cluster	Cluster 3				Cluster 4			
Distance	CC	CN	NC	NN	CC	CN	NC	NN
# samples	0	300	0	0	300	0	0	300

Table 2.7: *The two clusters of silence samples resulting from K-Means clustering with squared Euclidean distance.*

Cluster	Cluster 1				Cluster 2			
distance	CC	CN	NC	NN	CC	CN	NC	NN
# samples	8	300	8	299	292	0	292	1

whereas Hierarchical clustering generally fails, with few exceptions. Table 2.7 shows an example of K-Means clustering with Squared Euclidean distance.

One important outcome which emerged is the similarity between the results obtained for 7.5 and 15 ips datasets. In general, this result was expected, since the only differences are in the cut-off frequency in CCIR equalization (from 2 kHz to 4 kHz) [81] and this should not compromise the analysis. While the clusterings obtained from the white noise recordings were expected, the one obtained by the silence tracks can be explained with [98], where Mallinson analysis found that the dominant noise source in modern tape recorders is mostly originated from both the reproduce head and the recording medium itself and not from the write head. Therefore, in the case of silence samples, the background noise due to the write head is not powerful enough to be discerned from the one generated from the reading one [102].

Classification results

The aim of the following tests is to evaluate if Decision Tree, K-Nearest Neighbors, and Support Vector Machine (SVM) are able to discern:

1. the correct equalization and wrong equalization;

2. the correct equalization, CN, NC;

3. all four pairs of pre- and post-emphasis juxtaposition;

4. post-emphasis curves.

Table 2.8: *The classification performance using cubic SVM algorithm on white noise samples with four combinations of filters. The accuracy of this test is 0.97.*

filtersChain	Recall	Specificity	Precision
CC	0.907	0.996	0.986
NC	1	1	1
CN	1	1	1
NN	0.987	0.969	0.926

The last task derives from the results of the unsupervised learning tests. Initially, it was not considered because the main objective of this study is to detect post-emphasis equalization or at least understand incorrect equalization. Nevertheless, this information can be useful for digitized audio recording with no information about equalization.

The classification results confirm the previous analysis: datasets B and D (white noise) discern the correct equalization and the two wrong chain of filters, whereas the silence datasets detect only the post-emphasis curves. As for unsupervised learning tests there are no pronounced differences between 7.5 ips and 15 ips. In general, analyzing the first two groups of tests on the white noise datasets, the Accuracy, Recall and Specificity are 1, or very close. In both datasets, the best result is obtained with the Decision Tree classifiers (simpleTree, mediumTree, complex-Tree), where all the performance indexes are exactly 1.

Analyzing the results for the classification in four distinct pairs, for 15 ips samples performance indexes are equal or near to one for CN and NC, but not for CC and NN classes. In other words, the classifiers correctly recognize the wrong equalization pairs but have some difficulties to discern the correct pairs (CC, NN), confirming the results obtained with clustering analysis. For 7.5 ips, an unexpected result arises with cubicSVM on white noise samples dataset: the indexes are 1 for CN and NC classes and tend to the same value for the CC and NN classes. In other words, the classifier is able to recognize all the four type of samples. This fact could derive from non ideal analog filters or to small calibration misalignment of the tape recorder. More details are shown in Tab.2.8, where the Accuracy of the classification is 0.97.

The best result obtained in the last group of tests is with cubicSVM on the silence samples dataset. As just observed in the unsupervised learning experiments, the silence samples allow to precisely detect the post-emphasis equalization and this fact is prove with Accuracy, Recall, Specificity exactly 1.

2.4.2   Experiments with real datasets

Considering the excellent results of the previous experiment based on samples generated in laboratory, a further level of complexity has been tested using datasets extracted from real digitized audio recordings. Each sample is composed of the 13 Mel-Frequency Cepstral Coefficients (MFCCs). An evidence of the previous experiment is the small difference between 7.5 and 15 ips datasets. For this reason, only the first speed was tested in this new experiment. Some dif-

Table 2.9: *The number of samples for each real dataset.*

Dataset	CC	CN	NC	NN
A	74	73	67	69
B	221	221	263	263
C	0	0	40	40

ficulties on the collection of the samples derived from the necessity to have the 100% certainty of the correct equalization of the tapes. The samples were extracted from six recordings and originated by six CCIR (C) and four NAB (N) recordings. Two digitizations (with correct and wrong post-emphasis curves) of all tapes were analyzed in order to find parts with background noise only, as for the previous experiment.

Three different profiles of noise characterize the samples [117]:

A between -50 dB and -63 dB, consists on noises in the middle of a recording, i.e. silence between words in a speech;

B between -63 dB and -69 dB , are noises due to registration head without any specific input signal;

C between -69 dB and -72 dB, from sections of virgin tape.

Table 2.9 shows the distribution of the samples for each dataset.

Some tests showed that no pronounced difference was visible by varying the length of the sample in the range from 0.5 to 1 second. For this reason, the smaller value was used for each sample. The same machine learning techniques were tested:

- Unsupervised learning: Hierarchical and K-Means clustering;

- Supervised learning: Decision Tree (DT), K-Nearest Neighbors, and Support Vector Machine (SVM).

Clustering results

An important result emerges from the clustering analysis of both dataset A and B: the capability to discern the samples into two clusters based on the pre-emphasis curves. For dataset A, the best method is Hierarchical clustering. As shown in Table 2.10, the best combination of the parameters are the Euclidean distance and the complete linkage method. A first cluster contains 89% of samples with CCIR pre-equalization, while a second one contains 76% of samples with NAB pre-equalization.

For dataset B, the best result is obtained with K-means clustering. As shown in Table 2.11, the best distance metric is cosine, which have a first cluster with 84% of samples with CCIR pre-equalization and a second one with 79% of samples with NAB pre-equalization.

Distance	Cluster 1				Cluster 2			
	CC	CN	NC	NN	CC	CN	NC	NN
sqeuclidean	**97%**	**80%**	25%	21%	3%	21%	**74%**	**67%**

Table 2.10: *Two clusters resulting from Hierarchical clustering algorithm on the Type A dataset, with euclidean distance and complete linkage method.*

Distance	Cluster 1				Cluster 2			
	CC	CN	NC	NN	CC	CN	NC	NN
sqeuclidean	**87%**	**80%**	30%	13%	13%	20%	**70%**	**87%**

Table 2.11: *Two clusters resulting from the K-Means clustering algorithm on the Type B dataset, with cosine distance metric.*

The main difference between the two dataset is the number of metrics with which the clustering works. For dataset A, several parameters configurations returned good clustering, while for dataset B few distance metrics worked. The reason can be found in a stronger characterization of the writing noise in samples with more loudness.

Unfortunately, dataset C is composed by few samples of only NAB pre-equalization. This simplifies the model of detection both for clustering analysis and classification algorithms. However, the analysis shows an important result which is worth being dealt with in more depth in further works: around 100% of the samples are divided on the base of post-emphasis curve. This result was expected, since virgin tape is not recorded (and so the writing head does not add any kind of noise), but it could be an important indicator to analyze recordings already digitized by third parties which did not give any information of the used equalization.

Classification results

For dataset A, several classification methodologies confirm the results obtained by the clustering analysis. Some classifiers are able to discern the samples of dataset A on the base of pre-emphasis equalization. The best performance are obtained with Decision Tree (DT) and Support Vector Machine (SVM) and are shown in Table 2.12. Concerning the first technique, the accuracy is higher than 0.80 and recall near 0.95. Precision is slightly smaller (0.77), which means that

Algorithm	Type	Accuracy	Recall	Specificity	Precision
DT	Medium	0.83	0.94	0.7	0.77
SVM	Gaussian	0.80	0.89	0.7	0.76

Table 2.12: *The best performance of two classification algorithms trained for two classes that differs for pre-equalization for samples of Type A.*

Algorithm	Type	Accuracy	Recall	Specificity	Precision
SVM	Linear	0.7	0.8	0.57	0.70

Table 2.13: *The best performance of SVM algorithm trained for two classes that differs for pre-equalization for samples of Type B.*

23% of the samples are classified in a wrong class. Since this work is not meant to implement a tool that allows to modify the original digitization of the audio tape but only to check if it was digitized with the right equalization, we want to be sure to have the higher probability to select the right class for a sample in order to create as less false positive as possible. The results obtained from the classification go in this direction [117]. The performance for dataset B are less positive. With this kind of samples, it is more difficult to discriminate the pre-emphasis curve. However the trend follows the behavior observed in the clustering analysis. The best result is obtained with Support Vector Machine. The performance is presented in Table 2.13.

2.5   Stemmatics in tape music

As presented in Section 2.1, tape music is a genre in which the tape can be considered the final product of the creative process. In some cases, the composer did not provide a score. Therefore the carrier assumes a paramount role in the philological analysis. Many musicologists tackle the problem of stemmatics (or in Latin, *stemmata*) in their analysis of tape music works [150, 123], which consists on [144]:

- constructing the *stemma codicum* (recension, or the Latin *recensio*) by starting with a set of sources (all the different witnesses of that musical work), analyzed by comparison (the Latin *collatio*);

- selection (or *selectio*), where the original source is determined by examining variants, selecting the best ones [139].

This work presents an innovative approach to this philological analysis by adapting methodologies typically used in the field of forensic science: the *stemma codicum* is considered as an audio phylogenetic tree (APT) [143].

As described in the Section 2.2.1, editing of the tape was a widespread techniques. A wider analysis of the techniques adopted in tape music is presented in [60]. Some of them, not yet cited, concern the superimposition: a new signal can be recorded over the old one. It is also possible to record a new sound erasing the previous recordings or summing the two signals. Henceforth, these techniques are referred as *overdubbing*. The peculiarity of these alterations is the irreversibility: a new version of the opera (*witness* in philology) is created.

Digitization introduces an anomaly in the analysis which is not performed with the original analog documents, but with the digital ones. An analog tape can be digitized several times, creating similar digital versions of a unique document, which can differ from each other in the configuration of the tape recorder, in the quality of the equipment and in the format of the digital audio files. This study tackles only the first problem, concerning different configurations of the tape recording with the supplement of an external noise reduction system. As presented in Section 2.2.2, a tape could be recorded and read with six standard speeds. The wrong configuration of this parameter implies a time stretch and a pitch change that heavily distorts the audio signal. In some instances, such effect was deliberately used as a technique for altering the signal [60], but this aspect goes beyond this preliminary work. Section 2.2.2 describes in-depth the other main parameter to configure on tape recorder: equalization. The wrong choice implicates a non-flat frequency response applied to the audio signal. A further encoding that, if applied during the recording, needs to be compensated by the noise reduction system. If not compensated, the signal deeply change according to the type of system. The most common are Dolby A and Dolby SR (professional), Dolby B and Dolby C (domestic) and dbx Types I (professional) and II (domestic) [21]. In this work, the opposite case is also explored: the application of a noise reduction system during the digitization phase, even if this kind of system is not used in the recording of the tape. By only considering the combination of all these configurations, it is evident that many different digitized versions of the same tape can be obtained. Several other variables could extend this problem: different tape recorders, miscalibrations, damages or syndromes appearing during

the time, etc. This is only to indicate the complexity of this problem, but it goes far beyond this preliminary work, which seeks to prove the effectiveness of this new approach with a simplified model that nonetheless includes the main alterations.

The approach, developed with Simone Milani and Sebastiano Verde and the supervision of Sergio Canazza, is based on an algorithm created in the area of multimedia forensic. In this field, recent research investigates on new phylogenetic reconstruction strategies applicable on several multimedia contents (audio, video, image) [109, 103, 104]. An example is hereby given. A digital image can change over time (e.g resizing, applying filters, add text, etc.) and several versions can be generated, which can generate other versions as well [86]. These versions are defined as *near-duplicate* (ND). In case of images, the reconstruction of the generation process is named image phylogeny tree (IPT), and several reconstruction algorithms are proposed in literature [59, 57, 99]. Although a lot of work has been done for image and video, for audio only few algorithms are proposed in literature [109, 143].

From the phylogenetic point of view, the digitization represents a strong deviation from [104, 143]. The approach proposed in this work handles this aspect considering all the digitized versions of the same tape as a single node in the phylogenetic tree, creating a new level only if the tape is irreversibly changed. The proposed solution requires in input several audio files (witnesses) which are analyzed and grouped according to their shared characteristics. According to the algorithm, they are listed in a dependency tree in order to derive relationships between them. The approach is based on the analysis of the spectrograms of the digitized audio signals using computer vision techniques. The most likely transformation is inferred and characterized using a dissimilarity matrix. Such dissimilarity metrics are then used to characterize the edge weights of a complete graph where nodes correspond to the acquired audio tracks. By running a minimum spanning tree (MST) algorithm, it is possible to estimate the phylogenetic tree that links the different contents [143]. The algorithm described in the next section was tested in order to prove its robustness and accuracy by considering the main editing techniques and the most diffuse configurations of the tape recorder.

2.5.1 Algorithm

The proposed algorithm is based on techniques of computer vision applied to the spectrogram of an audio recording. Its input is the time-frequency representation of the audio signal, considered as in Figure 2.4. Through a feature extraction algorithm, a set of local spectral fingerprints were extracted in order to align pairs of images. After this alignment step, the pairs of images were compared in order to discover if any tape editing operation interlies between them. Similar strategies are described in [149, 146] for digital recordings, but for analog recordings, the digitization process adds several issues that required customized solutions.

The algorithm inputs are a set of N digitized audio recordings. For each couple of them, the dissimilarity is evaluated. The result is a dissimilarity $N \times N$ matrix $D = [d_{i,j}]$, where $d_{i,j}$ denotes the dissimilarity between the i-th and the j-th tracks. Based on this matrix, the algorithm builds a complete direct graph, where the nodes correspond to the analyzed set of tracks and edge weights are the computed dissimilarity values.

As shown in Figure 2.5, the methodology can be divided in the following steps [144]:

(a) *Original tape*

(b) *Modified version of the original tape*
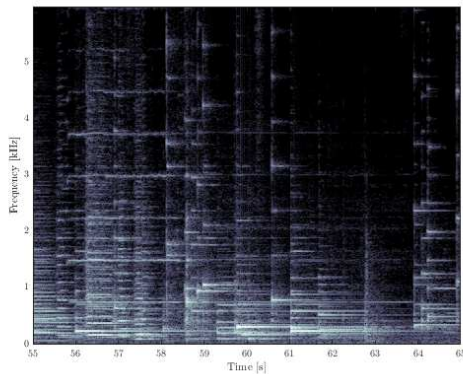
(c) *Spectrogram of the original tape*
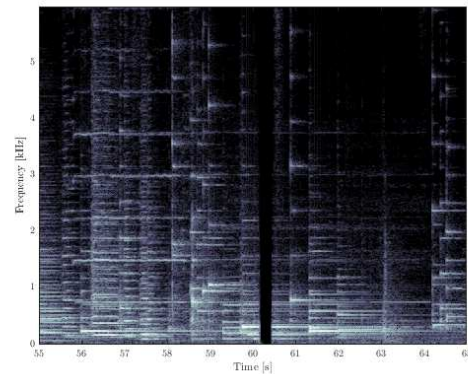
(d) *Spectrogram of the modified version of the tape*

Figure 2.4: *Example of near-duplicates (witnesses) [144]. In the middle of the tape (a), a piece of leader tape was added, thus obtaining the modified version (b). The difference between the two versions can be clearly observed by comparing the corresponding spectrograms (c) and (d).*

1. audio pre-processing;

2. leader tape detection;

3. spectrogram registration;

4. overdub detection;

5. estimation of the phylogenetic tree.

Audio pre-processing

As mentioned earlier, the spectrograms of all the audio signals x_i are extracted by using the well-known short-time Fourier transform. The chosen windowing techniques is the Hamming,
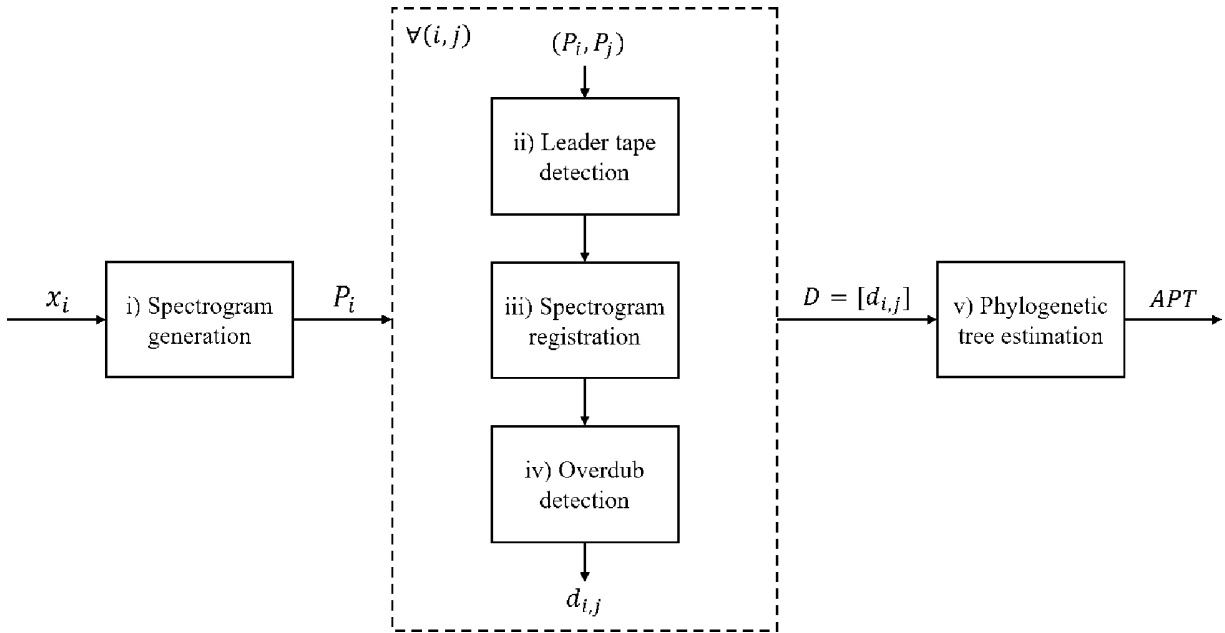
Figure 2.5: *Block diagram of the proposed algorithm [144]. Input consists of the digitalized audio tracks x_i, $i = 1, \ldots, N$, and output is the estimated audio phylogeny tree (APT).*

with a frame length of 4096 samples and an overlap rate of 75%. Considering the fact that most of the spectral information is found in the low frequencies range, the range of analysis has been reduced from 0 to 6 kHz, where 512 linearly-spaced frequency values were extracted (N_f). Each coefficient of the spectrogram is then associated to a greyscale pixel. The result is a $N_f \times M$ image $P_i(u, v)$, where M is the number of windows used in the previous step. the pixel intensity is obtained by converting the value $|X_i(f, m)|^2$ into an 8-bit integer. Furthermore, a reduction of the background noise was applied. The next step consists on the computation of the keypoints $\mathcal{K}_i = \{(u_k, v_k)\}$ from the image with the *speeded-up robust features* (SURF) algorithm [12].

Leader tape detection

As suggested in the section's title, in this phase a pair of spectrograms (P_i, P_j) is computed in order to detect if, in one of the two digitized recording, a leader tape was inserted. The observation at the base of the detector is that, whether a leader tape is present, there is no single affine transformation that maps the keypoints found on P_i on those of P_j. The reason is that the keypoints lying in the portion of spectrogram after a leader tape insertion will carry an offset in their time coordinates with respect to those found on a spectrogram that does not contain such insertion.

Here are the five points of the algorithm [144]:

1. Compare two sets of Keypoints $(\mathcal{K}_i, \mathcal{K}_j)$, in order to find a subset of marched pairs $(\mathcal{K}_i, \mathcal{K}_j)$ by comparing the related descriptor. Based on these matched pairs, the algorithm estimates the optimum geometric transform mapping P_i into P_j with the RANSAC algorithm [69].

As shown in Figure 2.6, if a leader tape is present, the set of inlier points returned by the algorithm will converge to a subset of keypoints belonging to only one of the two portions of the spectrogram separated by the leader.

(a) *SURF keypoints*   **(b)** *Inlier SURF keypoints after RANSAC*
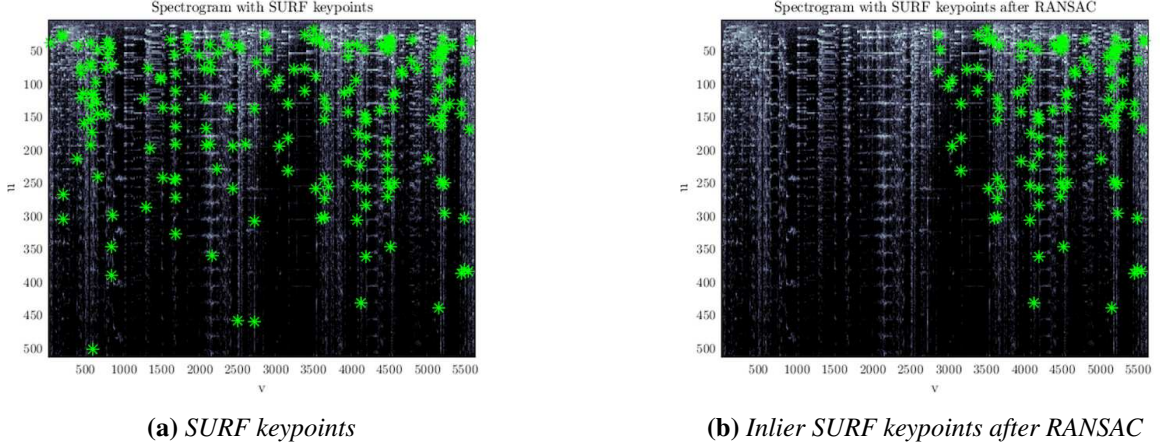
Figure 2.6: *Spectrogram image $P_i(u, v)$ of an audio track $x_i(n)$, with green asterisks representing the detected SURF keypoints. In 2.6b, it is possible to see the effect of the RANSAC algorithm: the inlier points left on the spectrogram are those located on the right of the leader tape [144].*

2. Define a function $g_i(v)$ counting the number of keypoints detected in $P_i(u, v)$ for each image column v. In order to avoid strong oscillations, this function is processed with a moving-averge LP filter. Then, define $g_i'(v)$ as the number of inlier points left on $P_i(u, v)$ after the RANSAC algorithm. In presence of a leader insertion, distance $|g_i(v) - g_i'(v)|$ shows an evident step that can be detected by looking for gradient peaks.

3. Let v_l be the coordinate associated to the detected step. Define the following sets:

$$\mathcal{K}_i^{(L)} = \left\{ (u_k, v_k) \in \mathcal{K}_i \,|\, v_k < v_l \right\},$$
$$\mathcal{K}_i^{(R)} = \left\{ (u_k, v_k) \in \mathcal{K}_i \,|\, v_k > v_l \right\}, \tag{2.3}$$

i.e., the subsets of keypoints found on the left side (L) and on the right side (R) of the spectrogram with respect to the candidate leader location. Similarly, define $\mathcal{K}_j^{(L)}$ and $\mathcal{K}_j^{(R)}$.

4. Compute a new geometric transform estimation, on the left and right portion of the images separately, according to the subdivision defined in (3). The estimated models are 3×3 homography matrices, $H^{(L)}$ and $H^{(R)}$, from which it is possible to extract the translation components along the v direction, $t^{(L)}$ and $t^{(R)}$. The length of the candidate leader is then given by

$$w_l = |t^{(L)} - t^{(R)}|. \tag{2.4}$$

If $w_l \neq 0$, the algorithm concludes that a leader tape is present within the current spectrogram pair.

The final part of the algorithm concerns the inferring of the correct phylogenetic relation that links P_i and P_j, namely whether P_j was derived from P_i by inserting a leader tape or vice versa. To achieve this, the average spectral energy around the detected location is measured, considering that leader tapes are characterized by a very low-energy region in the related part of the spectrogram. When a significant difference between the average energies, it is possible to conclude that the phylogenetic ancestor is the one related to the highest energy content. Therefore, if $\sum_u P_j(u, v_l) \gg \sum_u P_i(u, v_l)$, then P_j is assumed to be the phylogenetic ancestor. The dissimilarity matrix is updated with $d_{i,j} = +\infty$, therefore indicating that a phylogenetic relation from i to j is not possible, and the algorithm moves to the next pairs. Otherwise, if $\sum_u P_i(u, v_l) \ll \sum_u P_j(u, v_l)$ or $\sum_u P_i(u, v_l) \simeq \sum_u P_j(u, v_l)$, the algorithm continues with the next steps.

Spectrogram registration

The spectrogram registration consists in warping P_i towards P_j according to the geometric transform estimated through their matched keypoints. This step is necessary, since the following ones required aligned spectrograms. If a leader tape has been recognized in the image P_j, the algorithm adds black pixels in P_i in order to compensate it in P_i. The band of black pixel has a length w_l and is centered in v_l. The next step consists of extecuting RANSAC on all the keypoints in order to estimate the global geometric transform H. Then, P_i' is obtained by warping P_i towards P_j according to H. The last step consists of computing the dissimilarity value $d_{i,j}$ as the MSE of P_i' and P_j:

$$d_{i,j} = \frac{1}{U \cdot V} \sum_{u,v} |P_j(u,v) - P_i'(u,v)|^2, \tag{2.5}$$

where U and V are spectrograms height and width in pixels.

Overdub detection

The fourth step consists on the detect of any overdub. As first step, the absolute difference between P_i' and P_j is computed in order to find the residual spectrogram $P_r(u, v)$ (Figure 2.7a). Then, an energy content function ($e(v)$) of the residual spectrogram over the time is computed. In order to detect strong variations, the derivative of the previous function is computed and a outlier detector is applied. The latter consists of a 3 scaled median absolute deviation (MAD) from the median. The result is a set of points $\mathcal{O} = \{v_k\}$ (Figure 2.7b). This set is processed in order to find a candidate overdub $[v_1, v_2]$: the couple of points which maximizes the average energy ratio between the regions inside and outside those points.

$$(v_1, v_2) = \arg \max_{(v_a, v_b) \in \mathcal{O}^2} \frac{\mathbb{E}\left[e(v)\right]_{v_a < v < v_b}}{\mathbb{E}\left[e(v)\right]_{v < v_a \vee v > v_b}}, \tag{2.6}$$

where $\mathbb{E}\left[e(v)\right]_{\mathcal{I}}$ denotes the expectation of $e(v)$ for $v \in \mathcal{I}$.

In order to infer the phylogenetic relation, the energy is compared inside and outside the detected overdub, considering P_i' and P_j, instead of P_r. The last step consists on a scan of the

(a) *Residual spectrogram*
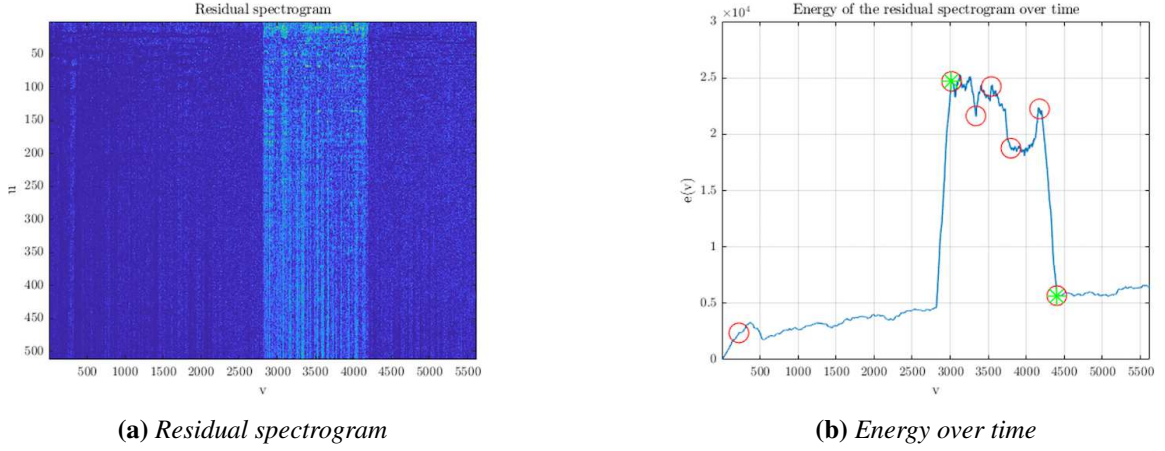
(b) *Energy over time*

Figure 2.7: *Residual spectrogram and related energy-over-time associated with a track pair (i, j) containing an overdub, which appears in 2.7a as a bright region with clean edges [144]. The red circles in 2.7b represent the detected outliers $v_k \in \mathcal{O}$ and the two points marked with green asterisks are the selected edges (v_1, v_2).*

spectrogram rows. For each row $(u = 1, \dots, U)$, the discrepancies between spectral energy inside and outside (v_1, v_2) are calculated in in P_i and P_j, as follows:

$$
\begin{aligned}
c_i(u) &= \left| \mathbb{E}\left[P_i\left(u, v\right)\right]_{v_1 < v < v_2} - \mathbb{E}\left[P_i\left(u, v\right)\right]_{v < v_1 \lor v > v_2} \right|, \\
c_j(u) &= \left| \mathbb{E}\left[P_j(u, v)\right]_{v_1 < v < v_2} - \mathbb{E}\left[P_j(u, v)\right]_{v < v_1 \lor v > v_2} \right|,
\end{aligned}
\tag{2.7}
$$

The spectrogram presenting a higher $c(u)$ for the majority of rows u is considered the phylogenetic descendant, in other words the overdubbed sample. Likewise, for leader detection, if j is chosen as ancestor of i, the algorithm sets $d_{i,j} = +\infty$. Otherwise, the dissimilarity value $d_{i,j}$ computed during the spectrogram registration is maintained.

Tree estimation

The result of the previous step is a dissimilarity matrix M, from which a undirected graph $\mathcal{G} = \{\mathcal{V}\mathcal{E}\}$ is computed. This graph has N nodes as the number of audio recordings, and each edge (i, j) exists if and only if $d_{i,j} < +\infty$ and $d_{j,i} < +\infty$. Then, a maximal clique algorithm is computed on the graph. The resulting $\mathcal{C}_1, \dots, \mathcal{C}_K \subseteq \mathcal{V}$ are used to calculate a $K \times K$ clique-dissimilarity matrix $D_\mathcal{C}$, as follow:

$$
D_\mathcal{C}(p, q) = \frac{1}{|\mathcal{C}_p||\mathcal{C}_q|} \sum_{i \in \mathcal{C}_p, j \in \mathcal{C}_q} d_{i,j},
\tag{2.8}
$$

where $|\cdot|$ denotes the cardinality of a clique.

From this matrix, a complete directed graph $\mathcal{G}_\mathcal{C} = \{\mathcal{V}_\mathcal{C}, \mathcal{E}_\mathcal{C}\}$ is created. Each node is a clique of the undirected graph $\mathcal{G}$ and each edge (p, q) has a weight equal to $D_\mathcal{C}$, corresponding to the average dissimilarity between the audio tracks belonging to the p-th and the q-th cliques.

The last operation is the computation of the phylogenetic tree $\hat{\mathcal{G}}_{\mathcal{C}} = \{\mathcal{V}_{\mathcal{C}}, \hat{\mathcal{E}}_{\mathcal{C}}\}$, which is the the directed rooted spanning tree with minimum weight (minimum spanning arborescence).

$$\hat{\mathcal{E}}_{\mathcal{C}} = \arg \min_{\mathcal{E}^s \subset \mathcal{E}_{\mathcal{C}}} \sum_{(p,q) \in \mathcal{E}^s} D_{\mathcal{C}}(p, q). \tag{2.9}$$

$\hat{\mathcal{G}}_{\mathcal{C}}$ is computed using the Chu-Liu/Edmonds' optimum branching algorithm [45, 62].

2.5.2 Dataset

In order to assess the algorithm illustrated in the previous section, ten electro-acoustic works were selected. For each track, a relevant sample of two minutes were extracted. Seven versions of each sample were created by applying different digitization setups and by editing the tape. These seven versions (near-duplicate) represent the input of the algorithm are due to reconstruct the phylogenetic tree.

These versions were created by applying the following procedure. The two-minutes original samples were extracted from the digital versions of the ten electro-acoustic works. Each sample was then recorded on a virgin tape. The configurations of the tape recorder during the recording are listed in Table 2.14.

Table 2.14: *Samples of electro-acoustic music recorded on the experimental tapes and the related configuration.*

Samples			Recording Parameters		
#	Composer - Title	Year(s)	Speed	Eq.	DBX
1	Luciano Berio - Différences	1958-59	7.5	CCIR	yes
2	Pierre Boulez - Dialogue de l'ombre double	1985	7.5	CCIR	yes
3	Brian Ferneyhough - Mnemosyne	1986	7.5	CCIR	no
4	Brian Ferneyhough - Mnemosyne	1986	15	CCIR	yes
5	Bruno Maderna - Continuo	1958	15	CCIR	no
6	Bruno Maderna - Dimensioni II - invenzione su...	1960	7.5	NAB	yes
7	Bruno Maderna - Notturno	1956	7.5	NAB	no
8	Luigi Nono - ...sofferte onde serene...	1976	15	NAB	yes
9	Studio NPS - Interferenze II	1965-1968	15	NAB	yes
10	Studio NPS - Ricerca 4	1965-1968	15	NAB	no

The first sample of each tree consists of the correct digitization of the recording and represents the root of the respective tree. In the phylogenetic tree, each edge represents an editing operation (transformation), whereas each node includes several versions of the same tape, digitized with different parameters configuration. The use of several versions on the same node differs from normal phylogenetic analysis, but it is required, since the tape is not modified but only read with a different parameter. Only the main editing operations (transformations) have been considered for evaluating the proposed algorithm:

- addition of a leader tape within the tape;

- overdub with silence or with another track;

- addition of a splice within the tape.

The last types of splices were created as for those ones in Section 2.3.1, cutting the tape at 90°and joining them again with a splice.

The professional open reel-to-reel tape recorder Studer A810 [135] was used in order to record and read the experimental tapes. This machine provides four recording/replay speeds: 30 ips, 15 ips, 7.5 ips and 3.75 ips. The equalizations supported by the machine are reported in Table 2.15, which shows the time constants of the curves for each speed. As shown in the table, the only standard equalization for 30 ips is the AES. A unique equalization is present also for 3.75 ips recordings (both CCIR and NAB use the same equalization). Furthermore, for some samples, an external noise reduction system DBX type I was used.

Table 2.15: *The equalization standards supported by Studer A810 described by their time constants. Source: [135]*

30 ips	15 ips	7.5 ips	3.75 ips
AES: $17.5/\infty$	**CCIR:** $35/\infty$	$70/\infty$	$90/3180$
AES: $17.5/\infty$	**NAB:** $50/3180$	$50/3180$	$90/3180$

The dataset includes multiple digitizations of the same tape, which were created by using such configuration parameters randomly. These multiple copies have to be included in the same node of the phylogenetic tree.

2.5.3   Results

The metrics used to validate the methodology are [144]:

1. accuracy of the leader tape detector;

2. accuracy of the overdub detector;

3. comparison of the estimated phylogenetic tree with the ground-truth.

The presented methodology was tested with the dataset described in the previous section. The first preformed experiments showed the problem analyzed in Section 2.3.1: the splices which join cuts are barely visible (or even invisible) in the spectrogram or at least easily confused with other spectral features. This approach, based on computer-vision techniques applied to the spectrogram, is highly affected by this issue. For this reason, ground-truth trees were re-designed by merging the clusters of nodes induced by a splice with their phylogenetic parents. The performance concerning the detection of the leader tape and overdub are presented in Table 2.16, in terms of probabilities of correct-detection and false-positive.

Leader		**Overdub**	
TP	FP	TP	FP
90.0%	0.0%	75.0%	3.3%

Table 2.16: *Correct-detection (TP) and false-positive (FP) probabilities for leader tapes and overdubs.*

A performance of 90% correct detections, with no false-positive, denotes a high reliability on the recognition of leader tape additions within the tape. The results concerning the detection of the overdub are lower, but still satisfactory. Here, two considerations could be done: in some instances, the overdub is not recognized as its spectral fingerprint is not sufficiently visible compared to the background noise or its interval limits are not sufficiently sharp in the difference of spectrograms. The other consideration refers to the presence or absence of DBX or equalization, which may produce artifacts that confuse the features related to overdub, thus obtaining false-positive.
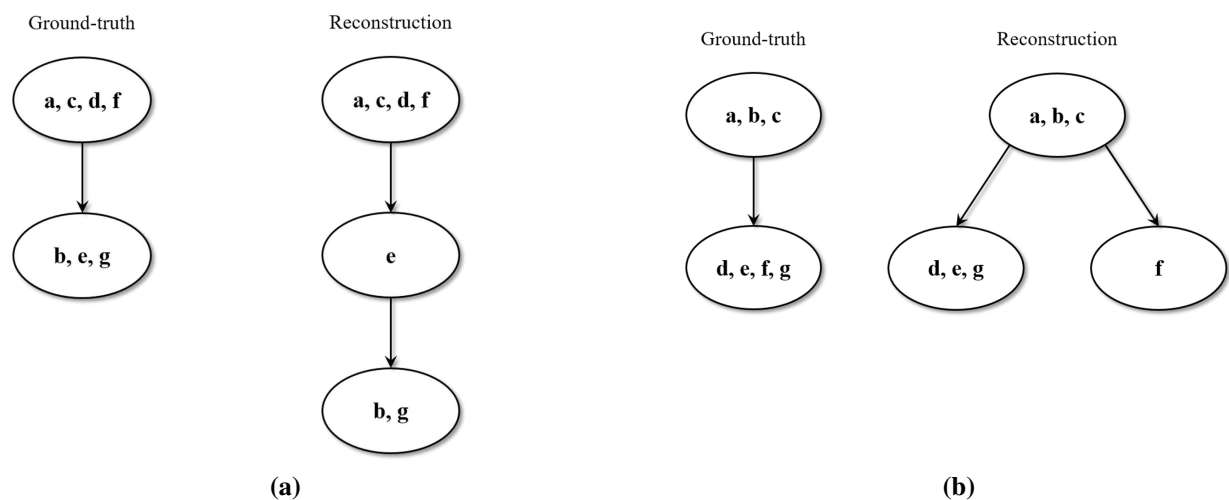
Figure 2.8: *Two examples of tree reconstruction containing over-clustering errors [144]. Datasets consist of seven audio tracks, $\{a, b, \ldots, g\}$.*

However, some errors or local noise can be tolerated, since the reconstruction process involves the whole dissimilarity matrix. Since the validation dataset is characterized by relatively small trees, the following results were obtained by qualitatively inspecting the estimated trees compared to the ground-truth. The main outcomes are:

1. In 50% of the cases, the estimated tree perfectly corresponds to the ground-truth.

2. In 40% of the cases, the estimated tree is not identical to the ground-truth but still makes sense in terms of phylogeny. For instance, in some instances it could be observed that certain cliques result over-clustered: tracks that should belong to the same meta-node are

split into several nodes, which can be siblings or parent-child relationships. However, the relative depths in the tree structure are maintained, and the overall phylogenetic sense is preserved. Figure 2.8 reports a couple examples of this scenario.

3. In 10% of the cases, the estimated tree shows some wrong phylogenetic relations (ancestor-descendant swaps) with respect to the ground-truth.

2.6   Experiencing audio documents

As mentioned in Section 1.2.1, the scientific debate concerning the preservation of historical audio documents often disregards essential aspects of music: the access, listening and experience. Audio players are mainly developed for digital music consumption and are not adequate for tape music. Nowadays, in most of sound archives, researchers can study digitized audio documents through Compact Discs or "iTunes-like" audio players. This kind of access is incomplete when the copy does not allow access to all the metadata and ancillary information [67]. In a broader sense, such limited access is unfaithful to the peculiarities belonging to the original (analog) recording. Sometimes this gap compromises the listening experience, such as for open works (see Section 2.1.1). In this case, the experience does not respect the original intention of the author, who based its work considering the replay devices. In the next section, the general objectives and methodology for a faithful access and interaction to the historical replay devices (suitable for open works) are presented.

Extending the concept in [18], the emulation of the tape recorder and carrier is not necessary and not recommended, since a lot of work is required for a provisional solution. The correct way to follow is the virtualization of the carriers and replay devices, that guarantees a faithful access and a durable solution [67]. Commercial software packages that reflect some characteristics of analog players (e.g., a gramophone) exist, but the virtualization is often incomplete and inaccurate.

A user interface (UI) that leverages on the appearance and behavior of physical artifacts is called *skeuomorphic*, which has proved beneficial in several cases. Some examples are [67]:

- the user is unable to adapt to a different, more abstract interface [80];

- realism enhances the user experience [132];

- the application purportedly aims at faithfully recreating the sense of interaction with an artifact of the past, which is no longer available in tangible form: in turn, this can be done for the pure pleasure of reviving the object [53, 87], or to obtain a milder learning curve by leveraging on the knowledge of the real object [4].

According to the last point, the touch screen of a mobile device is ideal to recreate the sense of interaction. The user can manipulate with her/his hands the virtualizations of the old replay machines and all their peculiarities and constraints for a more faithful experience. Furthermore, nowadays, the computational power is not a problem anymore, and even mobile devices are equipped with multi-core CPUs, augmented with specialized accelerators for 3D graphics and signal processing operations (e.g., multimedia decoding). This computational power, combined with sensors and a touch screen, provides an unprecedented multimedia and multi-sensory capabilities, supporting realistic interaction with digital objects. For this reasons, mobile devices are increasingly adopted in cultural heritage field, in particular for multimedia guides [145, 39], exploration of environments with augmented and virtual reality, and interaction with digitized artifacts. Several music applications have been developed [85], but no mobile app exists concerning the access and interaction with historical audio documents based on virtualization of

the replay machine. The first Android app that re-creates the experience of a reel-to-reel audio tape recorder was firstly presented in [36, 67], and was developed with the collaboration of Carlo Fantozzi and the supervision of Sergio Canazza. Section 2.6.2 describes its main features, assessment, as well as the new updates.

Recently, web technologies have also made a great step forward with the introduction of HTML5, WebGL and Web Audio API. The introduction of the latter enabled the creation of rich interactive audio applications without relying on the obsolete plugin such as Adobe Flash or Java [111]. The use of web applications provides numerous advantages. Firstly, any installation procedure is required with a significant impact on usability [147]. Furthermore, the portability across several platforms [119]. Some limitations exist (e.g. latency [147]) and some trade-offs have to be considered. However, browser can be considered an interesting platform for virtualization of an old replay device and for a complete musicological analysis. Three web apps designed respectively for the virtualization of a tape recorder and a gramphone and for the musicological analysis and segmentation of historical audio documents are presented in Sections 2.6.4 and 2.6.3.

2.6.1   Objectives for access and interaction

One of the main objectives for access and interaction is the *faithfulness* [36]. As for each single information related to the original audio document must be faithfully perpetuated, the presentation of such information must be as faithful to the original experience as possible [67]. According to [36, 67], the following three dimensions have to be considered in order to pursue this objective:

- faithful simulation of the experience of interacting with the playback device;

- faithful audio playback;

- faithful reproduction of all the metadata and contextual information.

In the following, these dimensions are detailed within the tape recorders domain.

Playback device

To reproduce the experience of operating with a tape recorder, the virtualization has to reproduce the original interaction. The aspects, buttons, switches, knobs, indicator lights and displays must be considered in order to recreate the experience. The visual feedbacks are also important, so displays and all moving parts must be updated in real time.

It is relatively easy to reproduce those elements where only the senses of sight and hearing are involved. The reproduction is more challenging for the elements that exert the sense of touch, prominently all the actuators: in this case, a touch screen is required. Mobile devices have an advantage because, with their affordable prices, they are endowed with touch screens.

Other features, such as quadraphonic audio are more easily implemented on desktop machines because of their superior computing power and the possibility to customize hardware

configurations (e.g. adding a professional audio interface). In general, a trade-off on the acceptable level of fidelity must be agreed upon during the design phase searching for the fairest balance between multimedia and multi-sensory capabilities, flexibility, and ease of implementation without incurring the complexity and costs of deploying supplementary custom hardware. A mobile device and a web app with these characteristics are presented respectively in Sections 2.6.2 and 2.6.3.

Audio playback

Three main features characterize the tape playback: replay speed, equalization and number of tracks. The first parameter is extremely important and concerns the quality of the tape, too. The higher is the replay speed, the higher is the quality of the recording. As reported in [30], both the noise spectrum and the high-frequency capability of the recording system increase in direct proportion to the tape speed. The six replay speeds are presented in 2.2.2. Only professional tape recorders have higher speeds, whereas the consumers ones only provide the slower. A wrong configuration of this parameter makes the listening impossible. Sometimes in a same tape, replay speed changes. It is clear how important the possibility to change this parameter is, which usually is not available in general audio players. Concerning the second parameter, a detailed study on equalization and its importance can be found in Section 2.2.2. In a tape, several tracks can be stored together. The number of tracks can vary from a minimum of one to a maximum of 24. Usually commercial devices and software provide only two output channels and only professional audio application are able to manage multi-track. It is clear how a CD or "iTunes-like" audio players are not adapted to experience these kinds of historical audio documents. In tape music, it is frequent to find quadraphonic works. A common approach is to mix two tracks in the left channel and the other two in the left one. For a faithful listening, this feature need to be managed. Furthermore, the possibility of selecting single tracks or a subset of them (as for some tape recorders) is an important functionality for scholars.

A remark concerning these first two points is however necessary: fidelity must be implemented in a sensible way. In some instances, where being overly faithful does not add to the interaction experience and it just makes the access to documents more involved. For instance, no reel-to-reel tape machine exists that plays all six standard replay speeds [21]. If the fidelity requirements are strictly followed, it is impossible to play all tapes with a single virtual interface. In this case, it may be more sensible to augment a knob so that all the replay speeds are available even if the actual device did not provide them. The same could be extended to equalization or other features [67].

Metadata and contextual information

As presented in Section 1.2.1, access copies need an adequate set of metadata and contextual information. While some metadata can be omitted, contextual information has to be provided as a whole. Photographic documentation of the carrier has to be included in order to report information about labels, edition boxes and other attachments, as well as clearly visible carrier corruptions. Furthermore, the video of the tape presented in the Section 2.2 has to be synchro-

nized with the audio signal in order to ensure the possibility to evaluate the impact and value of graphical marks, splices, corruptions, leader tape, etc., and therefore to discern intentional and unintentional alterations.

A group of features that deviate from the faithfulness issue concerns simplified access to discontinuities. A discontinuity detection algorithm (Section 2.3.2) can improve both the quality of a musicological research and the awareness on the importance of this information. In Section 2.6.4, a web application that includes the discontinuities extracted by the algorithm is presented.

2.6.2   Mobile-based virtualization: REMIND

As outlined in the introduction of Section 2.6, today's high-end mobile devices provide remarkable hardware specifications and the multi-sensory capabilities, with a relatively low price. For this reason, a mobile app that aims at re-creating the experience of a reel-to-reel audio tape recorder was created. Its name is REMIND (acronym for "Restoring the Experience: Mobile INterfaces for accessing Digitized recordings"). The code of this software is freely download-able[4]. As in a real tape recorder, it is possible to adjust parameters such as the speed and equalization, reproducing the behavior of a real machine. Furthermore, a complete set of metadata and contextual information is provided. In order to ensure a realistic interaction, the app is designed primarily for use on tablets with a 10-inch (or larger) screen.

Figure 2.9 provides an overall view of the app's structure. In addition to the main Android *activity*[5], three other activities provide the following functions:

Settings: gathers all preferences for the app and mainly dedicated to the management of the access copies;

Playlist: lists all the digitized audio document in the device and provides all the related metadata and contextual information;

Monitor Setup: allows the user to manage multi-track audio documents.

The first version of this app is described in detail in [36]. In the following sections, the second version is described, with a focus on the three dimensions and updates. Furthermore, in Section 2.6.2 the assessment of the app is reported.

---

[4]`github.com/CSCPadova/remind` (Retrieved 16, 2018)

[5]Android class in which is possible to place UI and manage the interaction with the user (`developer.android.com/reference/android/app/Activity`) (Retrieved September 16, 2018)
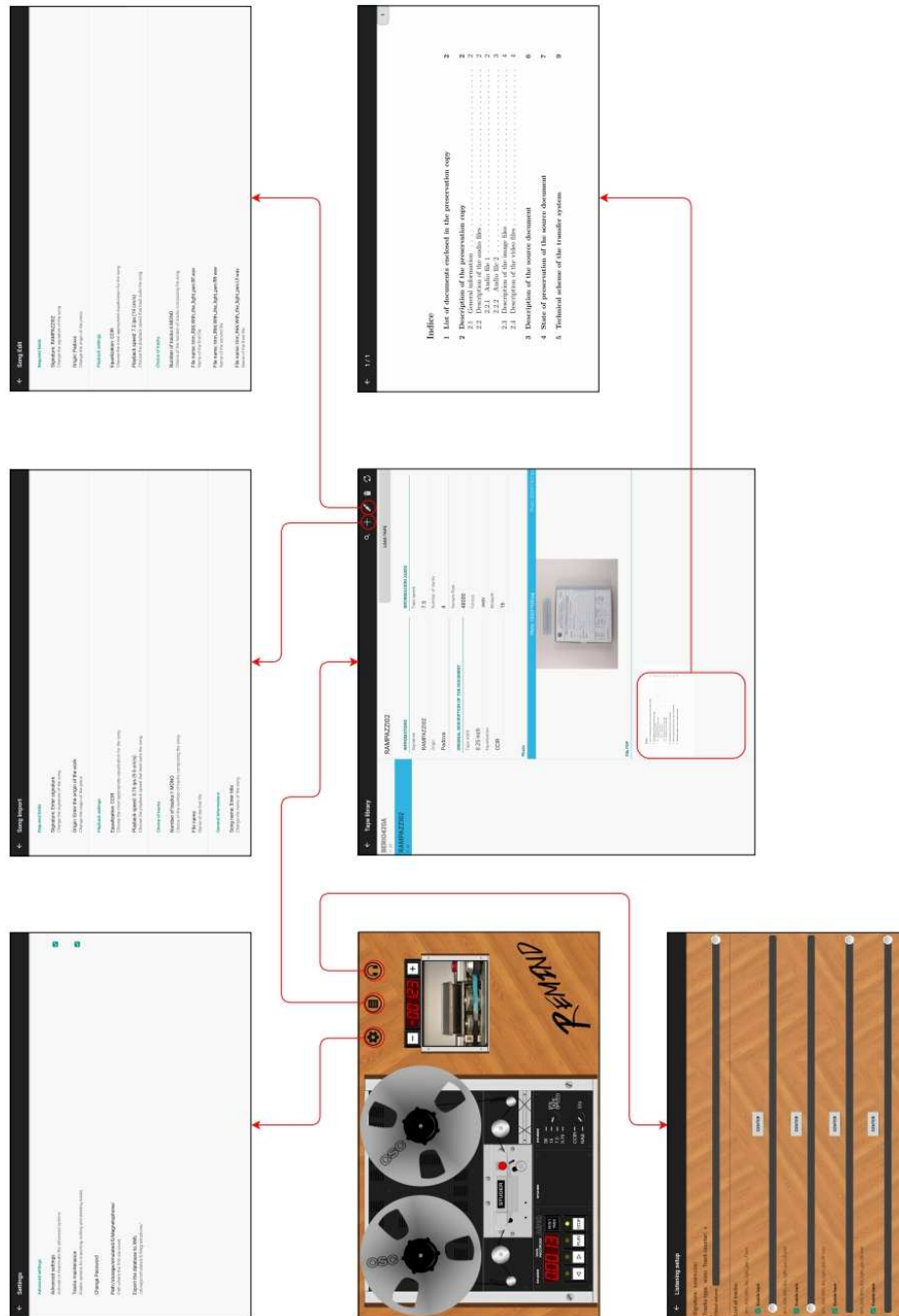
Figure 2.9: *The activities structure of the REMIND app.*

Figure 2.10: *The main activity of REMIND app [36]. A video of the original tape is shown outside the body of the tape recorder to improve its readability.*

Playback device

The main activity of REMIND app (Figure 2.10) provides a skeuomorphic representation of the Studer A810-VUK 2-track recorder (Figure 2.11). As it can be noticed, the overall appearance of the tape recorder are virtualized with touch buttons and knobs. The buttons provide the following functionalities: play, stop, fast forward, rewind and reset of the timer. The knobs allow to change replay speed and equalization. The recording functionalities of the original tape recorder were not implemented as they were deemed unnecessary to reach the faithfulness objective.

Several other features are represented, such as the movement of the tape with the rotation speed of each reel that is proportional to the amount of tape left in the reel and the timer. A complete description of reel rotation working principles is referred to [36]. The working principles concerning the video of the tape are discussed in Section 2.6.2.

The user interface (UI) is implemented with a custom Java subclass of the standard Android class *View*. Nearly all of the UI classes that appear in the main activity are custom: not only for the personalization of the appearance and behavior of the UI elements, but also to exploit hardware acceleration. The custom UI elements are 3D objects rendered into a *Canvas* allowing a consistently high frame rate for moving objects, and, at the same time minimizing the impact on the CPU. This choice also simplifies the scaling of the elements to support multiple display resolutions.

Audio playback

The most complex part of the REMIND is the audio backend. The app covers a significant part of the historical audio documents with a management of recordings formats from mono to

Figure 2.11: *A Studer A810 (VUK version) tape recorder [36]. In VUK versions, a separate console – not modeled by our app – houses the VU-meters, the channel control commands, and the monitor speaker.*

quadraphonic: one mono, one stereo, two mono and four mono files. Starting from the initial spatialization tool proposed in [36], a new track-management tool (Figure 2.12) was developed. In addition to overall volume, each track has a pan control which allows to allocate (or mix) the tracks in the desired output channel. Furthermore, a single track can be muted with a simple checkbox.

Android provides two main libraries to manage audio: OpenSL\ES[6] and the new AAudio[7]. Both are based and can be used by Native Development Kit (NDK), a framework based on C and C++. This kind of implementation provides high performance and low latency functionality. AAudio has been introduced in August 2017 and released for Android Oreo (API 27). Its main feature is the lightweight and the easiness of coding. However, the retro-compatibility is preserved with Oboe library[8]. This library is written in C++ and distributed by Google. Relying on the API of the device, Oboe automatically selects which library needs to be used: OpenSL ES from 16 to 26 API and AAudio from 27. This new version of REMIND has been re-implemented with Oboe.

The app manages NAB and CCIR equalizations and four replay speeds: 3.75, 7.5, 15 and 30 ips. In order to manage all the possible combinations of equalization and speed, the following chain has been implemented (Figure 2.13). The first step consists in removing the original equalization applied to the audio source. An inverse filter is convolved with the audio signal using an

---

[6]`developer.android.com/ndk/guides/audio/opensl/` (Retrieved September 16, 2018)

[7]`developer.android.com/ndk/guides/audio/aaudio/aaudio` (Retrieved September 16, 2018)

[8]`github.com/google/oboe` (Retrieved September 16, 2018)
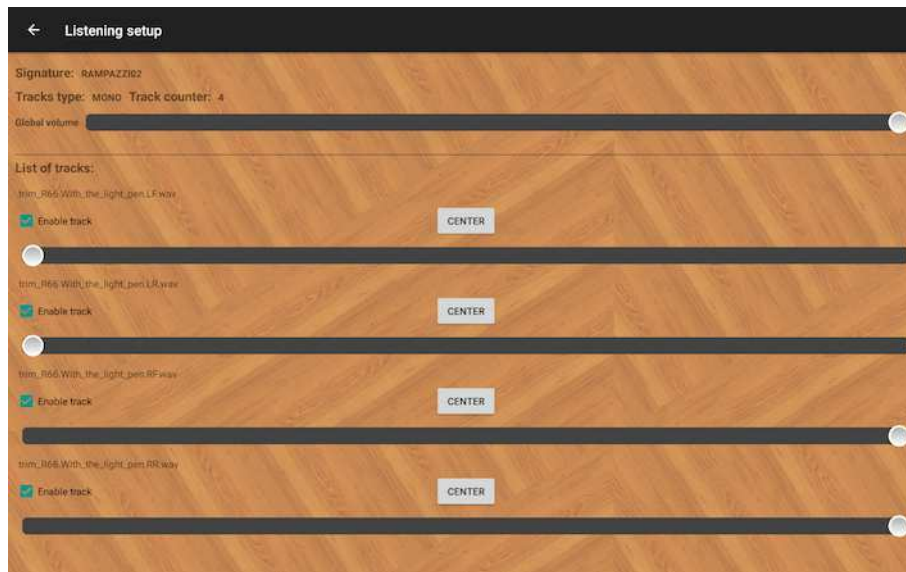
Figure 2.12: *The Monitor Setup activity of the REMIND app. Each track can be muted or panned.*
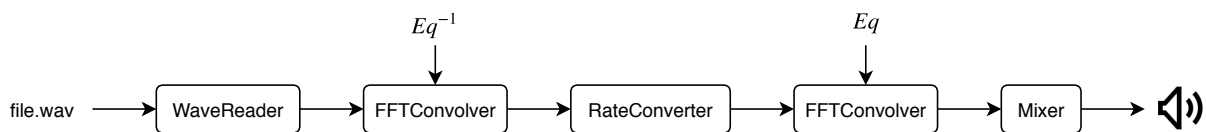

Figure 2.13: *Flowchart of the equalization chain.*

FFT algorithm. Then, the replay speed is changed in the node *RateConverter*. In the next step, the desired equalization is applied with another convolution with the related filter using again an FFT algorithm. In the last step, the audio signal is delivered to the *Mixer* which mixes and routes the signals to the output channels.

Metadata and contextual information

As stated in Section 1.2.1, the role of the metadata and contextual information is paramount. REMIND provides a complete set of metadata and information as well as the photos and video stored in the preservation copy.

The latter is immediately visible in the main activity. It is synchronized with audio, even when the replay speed changes. The frames are rendered at an appropriate rate to match the actual speed of the tape, even if it is different from the speed at which the video was captured [36]. The only quality reduction happens during reverse and fast forward, when it has been deemed acceptable to show only intra (key) frames during reverse and fast reverse. This decision aims to reduce the computational load of this operation. In the new version of REMIND, this feature has been enriched with two offset buttons. In order to correct the small misalignment between audio
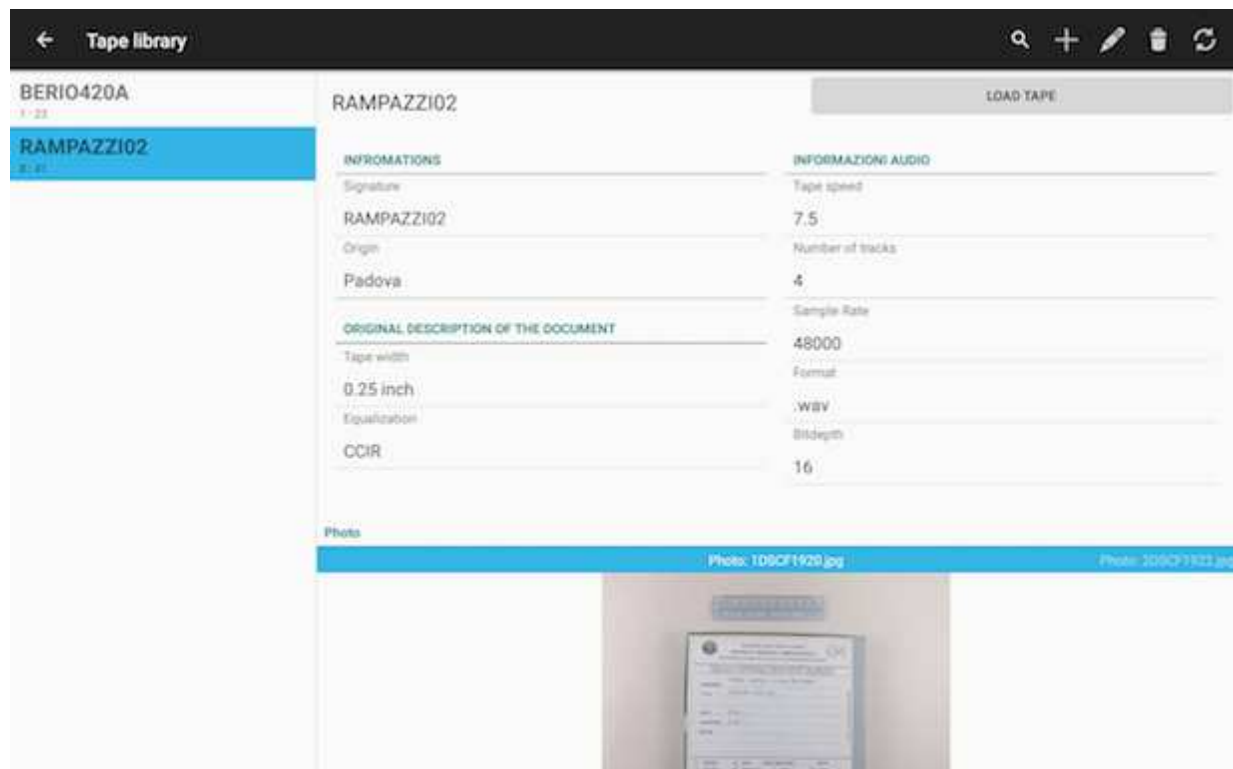
Figure 2.14: *The Playlist activity of the REMIND app. It contains a list of the access copies and their metadata and contextual information.*

and video, the buttons allow to increase or decrease the offset with a milliseconds precision. A panel, similar to the timer of the magnetic tape, indicates the offset. Furthermore, the buttons allow to exponentially accelerate the increase or decrease of the offset through a long pressure of the related buttons.

From the main UI, it is possible to enter into the Playlist activity through a simple icon. As it can be seen in Figure 2.14, the left part of the interface is characterized by a list of the audio documents available in the device. Each access copy can store audio files, video, photos and a pdf file. All the files paths and a basic set of metadata are stored in a SQLite database for Android[9], importable and exportable through XML file. Another way to populate the database is a wizard where all the basic metadata can be manually inserted and the files imported from the internal memory of the device. In order to protect the integrity of database and access copies, the operations of creation and update are allowed only after inserting the password.

After selecting the desired audio document, a multi-pane layout in the right part of the interface shows a basic set of metadata, a slideshow with the photos and a preview of pdf file with all the information related of the original audio document. The photos can be enlarged with a simple click. The possibility to load and read the pdf file originated during the creation of the preservation copy is a new feature of the last version. The Android basic library PdfRenderer.21

---

[9]`developer.android.com/training/data-storage/sqlite` (Retrieved September 16, 2018)

Table 2.17: *Experience of the professionals recruited for the assessment of the REMIND app [67].*

Experience	Professionals	Count
Up to 5 years	P1, P2	2
From 5 to 10 years	P3, P18, P19	3
From 10 to 15 years	P20	1
From 15 to 20 years	P4	1
From 20 to 25 years	None	0
From 25 to 30 years	P9, P14, P15	3
From 30 to 35 years	P5, P10, P16	3
From 35 to 40 years	P11, P12	2
From 40 to 45 years	P6, P13	2
More than 45 years	P7, P8, P17	3

was used for the file preview, but, considering its limited capabilities, it has been discarded for the pdf visualization. Android PdfViewer has been selected for its integrated features such as gestures, zoom, scroll, etc.

The feature presented in [67] concerning the intelligent assisted access to discontinuities was temporary removed in the new version. The idea is to develop an integrated system that can help scholars in the musicological analysis of the audio documents with advanced features such as the interface presented in Section 2.6.4.

Assessment

The usefulness of the REMIND app was assessed with 20 music professionals (henceforth indicated with labels from P1 to P20) of different ages, experiences and competences. Half of the subjects are musicologists, whereas the other ones are composers of electronic music (Table 2.18). As can be observed in Table 2.17, the 65% of the subjects have more than 25 years of professional activity, and a 25% have more than 40 years of experience in the field. As a consequence, the median age is high, over 50 years old (Table 2.19). The main tools exploited in their work are presented in Table 2.20 The main difference between the subject is that, on average, composers and young use more software tools respect to musicologists which rely on vinyls, CDs, scores or other artifacts.

After a complete explanation of the main functions of REMIND, the professionals were free to interact with the app running on a Galaxy Tab S (10.5 inch tablet). Then, they answer to a set of 14 questions related their experience with a grade in a 0 ("no", "strongly disagree", "totally lacking", "completely inadequate", "completely useless") to 4 scale ("yes", "fully agree", "nothing is missing", "perfectly adequate", "very useful"). Furthermore, in some questions a non-numerical comment was encouraged. The full version of the questionnaire and all the numerical answers can be found in Appendix B. A summary is presented in Table 2.21.

Table 2.18: *Main occupation of the 20 professionals recruited for the assessment of the REMIND app [67].*

Main occupation	Professionals
Composer of electronic music	P1–P8
Musicologist, instructor (conservatory)	P9–P13
Musicologist, instructor (university)	P14–P17
Musicologist, Ph.D. in musicology	P18–P20

Table 2.19: *Age of the professionals [67].*

Age	Professionals	Count
Below 30	P1, P2	2
From 30 to 50	P3, P4, P18–P20	5
From 50 to 70	P5, P6, P9, P10–P16	10
More than 70	P7, P8, P17	3

Table 2.20: *The tools used by the professionals involved in the study [67].*

Age	Professionals	Count
Max	P1–P5, P9, P12, P18, P20	9
Pure Data	P6	1
Audio editor	P1–P3, P5, P6, P9–P16, P18–P20	16
Music composing software	P1–P3, P5, P6	5
Open source software	P18, P19	2
Custom software	P15, P20	2
Listening to CDs	P5, P7–P16, P18, P20	13
Listening to vinyls	P17	1
Computer	P8	1
Scores	P5, P7, P8, P11, P13-P17	9
Study of the tools of the composer	P13	1

Table 2.21: *Summary of the results from the assessment of the REMIND app. Ratings were given on a 5-level scale from 0 (most negative) to 4 (most positive).*

	MEAN	MED	MIN	MAX
Q1: Does the app always present information in a clear fashion?	3.7	4	3	4
Q2: Are text messages and icons clear and unambiguous?	3.4	3	3	4
Q3: Is the navigation structure easy to remember?	3.8	4	3	4
Q4: Is the app sufficiently self-explanatory?	3.5	3	3	4
Q5: Is the touch screen a convenient mean of interaction?	3.7	4	3	4
Q6: Is the skeuomorphic interface actually interesting?	3.2	3	1	4
Q7: Is it correct to have sequential access as the only option?	0.3	0	0	1
Q8: Is the physical size of the tablet adequate?	3.8	4	2	4
Q9: Were all tasks as easy as expected?	4.0	4	4	4
Q10: Are metadata complete, with no useful metadata missing?	1.8	1	1	4
Q11: Is the app complete, with no useful function missing?	1.6	1	0	4
Q12: Are all implemented features actually useful?	4.0	4	4	4
Q13: Would the app allow you to work faster?	1.8	2	0	3
Q14: Would the app allow you to work more accurately?	3.2	3	1	4

The questionnaire was designed to investigate whether the app design was perceived as appropriate and met the needs of the professionals. In general, all the interviewed professionals provided positive feedbacks, even if the youngest (probably more accustomed to apps) tended to be a bit less captivated. From the average results, the app is judged clear (Q1, Q2), self-explanatory (Q4), easy to navigate (Q3) and any task was found unexpectedly difficult (Q9). A tablet was considered the right device for the app (Q8), and the touch screen a comfortable mean of interaction (Q5). Only two professionals stated that it would be better to have a larger touch display. All the features implemented in the app were consistently considered as useful (Q12). The skeuomorphic interface was generally credited as an interesting plus (Q6). Nonetheless, the professionals were eager to compromise fidelity for comfort [67]: nearly all of them asked (Q7) for philologically questionable means of accessing music. Besides random access, the professionals asked for what the authors call "assisted access", and chiefly for further automatic means of searching noises/defects, sounds, or user-defined sound structures, even across several pieces of music. Concerning features, the most requested additions were the possibility of analyzing the waveform and spectrum of the signal (spectrogram, periodogram), and the possibility of adding notes and bookmarks. These kinds of features are currently under development. In Section 2.6.4, a similar solution with features to navigate discontinuities and to annotate the analysis are integrated. In Q10 (metadata) and Q11 (features), the professionals also demanded additional metadata and features not presented in the first version. For this reason, the new version of REMIND app includes a pdf file containing all the metadata and the contextual information extracted and stored in the preservation copy. Regardless of their severe judgement for Q10 and

Q11, an overwhelming majority of the interviewed professionals declared that REMIND already has the potential of making their work more accurate (Q14), albeit not faster (Q13). The addition of a non-sequential access, albeit technically easy, requires to be included in a philologically-aware context and so a deep meditation on the design of the user interface.

2.6.3 Web-based virtualization: REWIND

As mentioned in the introduction of this Section, new web technologies such as HTML5 and Web Audio API increased the capabilities of a web browser, this being now also suitable for complex audio applications. Two applications based on these technologies are presented in this section as part of the same project: REWIND (acronym for "Restoring the Experience: Web Interfaces for accessiNg Digitized recordings"). The firstly developed interface was a philologically-aware virtualization of a gramophone for 78 rpm[10] discs [116]. Then a new interface which virtualize a tape recorder was created [67]. As for the REMIND app of the previous section, the replay machine is inspired to a Studer A810 (Figure 2.11). Figure 2.15 shows the two main interfaces of these web applications, which are freely downloadable[11][12].

Although the gramophone is not pertinent to tape music, it shares some peculiarities and problematics derived by old analog technology. For example, the "equalization problem" is taken to extremes. Before the introduction of RIAA (Record Industry Association of America) as global standard (1955), several standards were introduced by record companies [72, 21]. As result, several equalizations and speed standards complicate the digitization and the listening of these recordings. For this reason, an *ad hoc* solution was developed, which differs from the solution adopted for tape recorder virtualization, but can be considered as a valid alternative. As for REMIND app, both web applications were improved compared to the versions detailed in [116] and [67] respectively. To follow, the new functionalities and the two approaches to the "equalization problem" are presented.

The two web applications use the following web technologies: PHP, HTML5, CSS3 and relies on other external libraries such as jQuery[13]. The code has been updated in order to operate with the new versions of these technologies. The new versions support mp3, WAV and FLAC audio files as well as mp4 and WebM video files. Both applications have a new track loader based on SQLite that replaces MySQL, in order to be usable as standalone. A single digitized tape or disc can be inserted manually using the interface presented in Figure 2.16a or imported from a tracklist in JSON format. The list of songs can also be exported (in JSON format), as well as all the audio files, downloadable in a single ZIP file (Figure 2.16b). Moreover, new commands to upload and delete the audio and video files are introduced, combined with a set of control functions to verify the presence of the files after an upload (Figure 2.17). The set of metadata is still minimal and could be further improved to support all the metadata of a preservation copy.

---

[10]RPM is acronym of Revolutions Per Minute

[11]REWIND - Gramophone: `github.com/CSCPadova/rewind_gramophone` (Retrieved September 16, 2018)

[12]REWIND - Tape recorder: `github.com/CSCPadova/rewind_taperecorder` (Retrieved September 16, 2018)

[13]`api.jquery.com/` (Retrieved September 16, 2018)

Figure 2.15: *The main UI of the REWIND interfaces: the virtualization of a gramophone (a) and a tape recorder (b).*

However, the main attributes, such as equalization and speed replay, are included.

As for REMIND app, the skeuomorphic interface of the tape recorder is designed to support speeds and equalizations of a Studer A810. Two clickable knobs allow to select both parameters. The rotation mechanism has been sacrificed in order to improve the ergonomics, usability and compatibility of the interface. The interaction is usually performed with a mouse, thus simple buttons are provided to rotate the knobs. As for REMIND, the video of the tape is provided beside the tape recorder, and synchronized with the audio. In this application, the possibility to apply an offset is not yet implemented. Web Audio API is based on *AudioContext*, an audio graph composed by several nodes from the sources to an *AudioDestinationNode* which is the last node that routes the signal to real-time output device [1]. The equalization for this app is implemented with a *ConvolverNode* which convolves in real-time the signal from an *AudioBufferSourceNode* [1] (the node which streams the audio file) with an impulse response, preloaded during the initialization of the application. The idea is to use a different impulse response for each combination of speed-equalization couples, as for the mobile version. At the moment, the creation and the validation of these impulse responses is currently work in progress, but the software is ready.

The solution adopted in the gramophone app is structured in a different way. The skeuomorphic interface of the gramophone includes the possibility to select through a cursor the replay speed between 70 and 80 rpm, but it is also possible to pick it in a list of standards, in the panels below. The list includes these standards: Columbia (70 rpm), Victor e HMV (71.29 rpm), Acoustic Victor (76.59 rpm), Electronic Recording (78.26 rpm) e Acoustic Columbia e Vertical Recording (80 rpm). The equalization can be selected in the panels only. One panel is related to the equalization used during the digitization process, whereas in a second one a new one can be selected (ideally the original one). Some of the main standards are provided (RIAA, RCA, HMV, FFRR, NAB), but by simply configuring three parameters, a customized curve can be created. Usually equalization curves for gramophone are stated with a bass turnover (the cutoff

(a)

(b)

Figure 2.16: *Panels of the REWIND applications to import and export single files or list of audio documents.*

frequency of the filter applied to low frequencies) and a rolloff. The latter consists of the value in dB at 10 kHz. Considering the slope of the curve (6 dB/octave), it is possible to have the treble turnover of the filter at medium and high frequency. A further parameter is related to a Shelving filter of the very-low frequency, which is present in few standards. The customized curves are created by setting three range tags related to these parameters.

A set of 31 low-order filters, implemented with just as many *peaking BiquadFilterNode* [1], Their distance is one-third octave and thus their Q-factor is 4.32. The other parameters of each filter are the gain value and the central frequency. In order to avoid distortions of the signal, a small offset was added to limit the gain of a filter at +10 dB. The new version of the web application also includes the visualization of the selected curves. Three curves can be selected:

- the inverse curve, necessary to compensate the (incorrect) one applied during the digitiza-

TRACK LOADER			∧

LOAD TRACK		LUCIANO BERIO - BERIO TAPE 001 (1960)	🗑

		SELECT VIDEO TRACK (MP4, WEBM):	
AUDIO FILE: BERIO001.MP3	DELETE AUDIO	Sfoglia... Nessun file selezionato. Upload Video	

Year	Speed	Equalization	Video
1960	15	ccir	No

BERIO002.MP3 NOT FOUND		LUCIANO BERIO - BERIO TAPE 002 (1961)	🗑

SELECT AUDIO TRACK (MP3, FLAC, WAV):		SELECT VIDEO TRACK (MP4, WEBM):	
Sfoglia... Nessun file selezionato. Upload Track		Sfoglia... Nessun file selezionato. Upload Video	

Year	Speed	Equalization	Video
1961	7.5	ccir	No

Figure 2.17: *Panel of the track loader of the REWIND tape recorder. If a file is not found can be directly selected in this section.*

CORRECTIVE CURVE	READING^(-1)	READING	SHOWING CORRECTIVE CURVE: RIAA^-1 + NAB
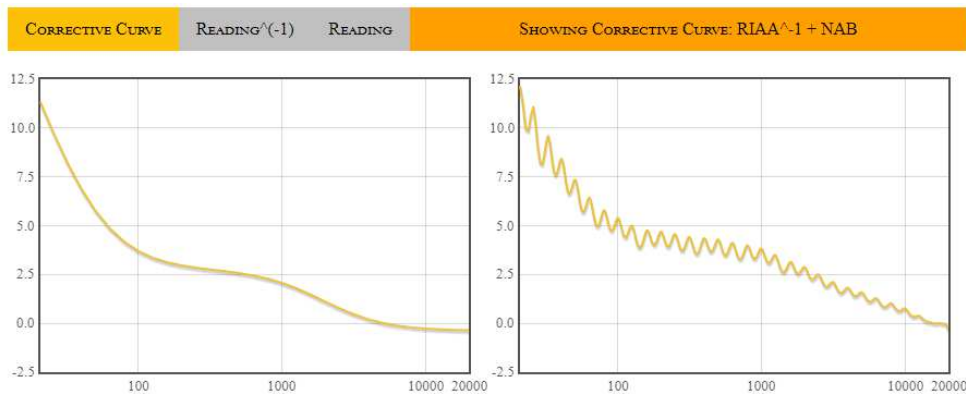
Figure 2.18: *Plot of the theoretical and real equalization curves applied to a virtual disk in REWIND gramophone interface.*

tion;

- the new curve, ideally applied to compensate the original equalization applied during the disc recording;

- the resulting curve merging the two equalizations.

The equalization plot is splitted in two graphs (Figure 2.18): the theoretical curve and the real one obtained by summing all the 31 filters. The wave shape of the signal is reduced by calculating the parameters of the filters (gain and Q factor) as proposed in [142]. The overall result is not completely smooth, but it follows the original shape. In case of big attenuations or boosts, the shape near the frequency boundaries are more irregular. A further improvement will be the addition of Shelving filter to reduce this behaviour. This feature is developed using Flot[14], a

---

[14]`flotcharts.org/` (Retrieved September 16, 2018)

javascript library for jQuery. The current limit of this compensation is the rotation speed. In this version, the change of speed in not yet considered in the estimation of the equalization curve.

2.6.4   Access Kit

The CSC procedure for creating preservation copy is supported by PSKit [22]. An operator inserts and uploads all the data, metadata and contextual information in this application, which creates the final preservation copy. This software is not studied for the access and the consumption of the audio document. However, the database could be exploited in order to fully exploit the complete preservation copy. In an open project with an international archive, a simple web interface has been developed at CSC with the contribution of the author to annotate the musicological analysis, but that does not provide full access to the data, metadata and contextual information. By extending this interface and the database of PSKit, a new web application was created and named *Access Kit*. Currently, this software is not yet completely stable, but it is a working proof of concept with an implementation of all the main features.

The architecture of this version of Access Kit is based on REST (REpresentational State Transfer) [19] and HATEOS (Hypermedia as the Engine of Application State)[15] principles. The server part was developed with Spring framework[16], whereas the client part with Angular JS 1.6[17]. The database is based on MySQL as for the PSKit.
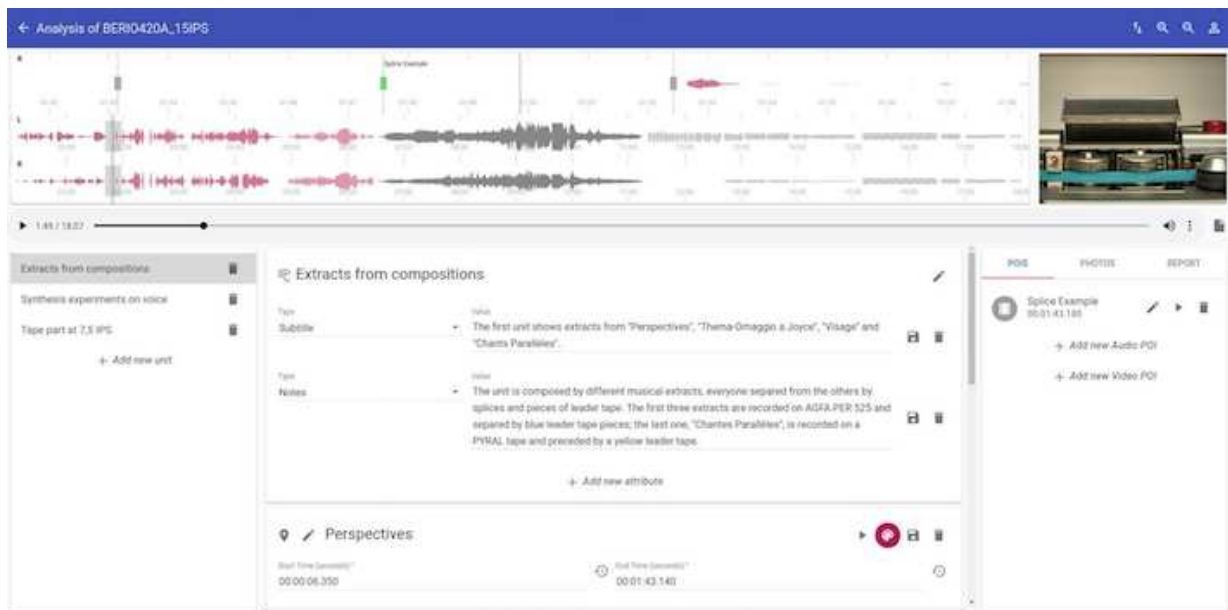
Figure 2.19: *The main interface of Access Kit.*

The main view of Access Kit is proposed in Figure 2.19. The interface can be divided in three parts:

- the audio/video player;

- the annotation section;

---

[15]`spring.io/understanding/HATEOAS` (Retrieved September 17, 2018)
[16]`spring.io/` (Retrieved September 17, 2018)
[17]`angularjs.org/` (Retrieved September 17, 2018)

- the contextual information panels.

The audio player has in turn three main features. The player bar has the main functionalities to play and stop the audio document and a bar to navigate audio content. As for REWIND, the video of the tape flowing on the head of the tape recorder is synchronized with audio. The last element is the waveform of the audio signal. It is based on a customized version of the library Peaks.js[18], developed by BBC research team. This library shows two waveforms[19]: one is the overall waveform whereas the other is a zoomed interval of the previous one. Furthermore, the user can increment or decrease the zoom interval. The new functionality implemented in Access Kit consists in a separate visualization the waveforms of the two distinct channels (if stereo), and the possibility to select the channel to zoom. Other native features integrated and extended in the application are the possibility to create colored segments and markers. These features are linked with the second and the third parts, respectively.

The annotation section recalls the first aforementioned interface and includes the possibility to subdivide the document in *units*. Each unit contains a set of descriptors, selectable by a drop-down lists and a set of events. Each event is characterized by a starting and ending timestamps as well as a text area, where a scholar can take annotations. The unit boundaries are the minimum and maximum values in the events timestamps. Each event is highlighted in the waveform with a color (which can be selectable). Only the events of a single unit at time are painted in the waveform.

The third part is divided in three panels. The first one is related to Points of Interest (POIs) which consist on instants when something significant happened. This feature is strictly connected with the concept of discontinuity. A POI is characterized by a timestamp and a description. If the POI is a discontinuity extracted with the method described in Section 2.3.2, it is possible to enlarge the extracted frame. As it can be noticed in Figure 2.20a, the pop-up also provides an evaluation of the POI. This feature is thought for a further development, where such information could help to discriminate false positive discontinuities. A second panel provides the thumbnails of the photos included in the preservation copy (Figure 2.20b). With a simple click, it is possible to enlarge the photos to the original size. The last panel allows to download a pdf file that contains all the metadata of the audio document. The last important feature is the exporting tool that produces a latex source file of the analysis, that can be easily customized before the creation of the final pdf.

Both the old and the new interfaces are useful tools for the segmentation process and can also be used to help the creation of access copies or segmented ones. For the first one, a cutting tool has been developed. The simple application presented in Figure 2.21 allows to cut audio files considering events or units. This application is written in Java and exploits Sound eXchange (SOX) library[20]. A further development is the integration of this feature in Access Kit without an external application.

---

[18]`bbc.co.uk/rd/blog/2013-10-audio-waveforms` (Retrieved September 17, 2018)

[19]The waveforms are computed with the tool *audiowaveform*, developed by BBC, `github.com/bbc/audiowaveform` (Retrieved September 17, 2018)

[20]`sox.sourceforge.net/` (Retrieved September 17, 2018)

(a)   (b)
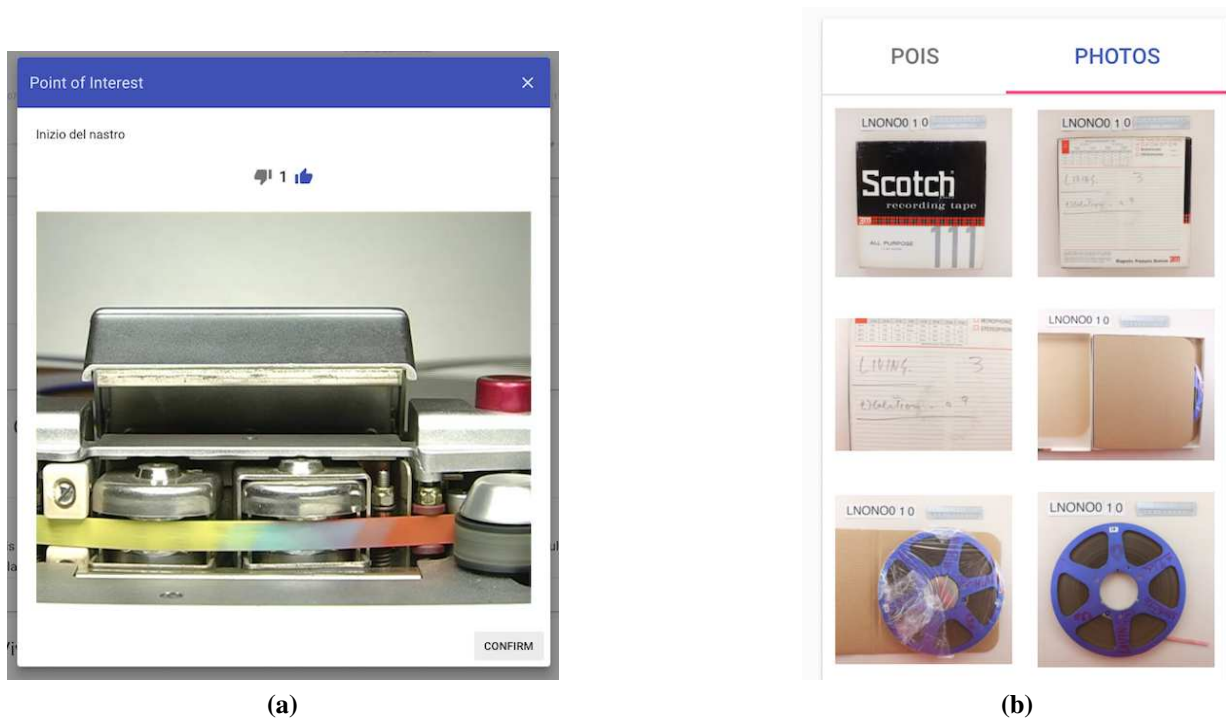
Figure 2.20: *Points Of Interest (POI) pop-up with a frame of a splice 2.20a and a panel with the thumbnails of the photos included in the preservation copy 2.20b.*
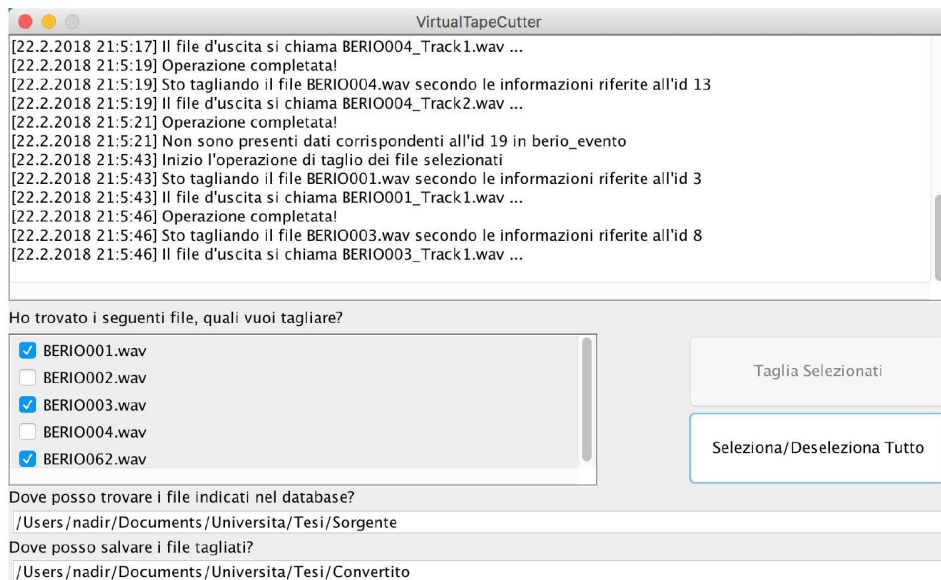
Figure 2.21: *Application for the segmentation of audio files on the basis of units or events timestamps provided by a musicological analysis.*

2.7 Case study: Arab-Andalusian corpus

Computational Musicology and Music Information Retrieval (MIR) required solid corpus for the implementation of data-driven research tasks [129]. The aim of the ERC research project Comp-Music was to advance the automatic description of music by emphasizing cultural specificity and by carrying research within the field of music information processing with a domain knowledge approach. During this project five corpus concerning non-western musical tradition were created: Hindustani (North India), Carnatic (South India), Turkish-makam (Turkey), Beijing Opera (China) and Arab-Andalusian (Morocco).

This work focuses on the latter one. The creation of the Arab-Andalusian corpus started in 2013, at the beginning of the CompMusic project. The challenges tackled to create a complete research corpus as well as the origin of the recordings are described in [131]. According to the CompMusic principles, the corpus is created respecting five main criteria, formalized in [128]: purpose, coverage, completeness, quality and re-usability. The general purpose of this corpus consists in pushing forward the state of the art in music information retrieval and computational musicology, by addressing the research challenges that this tradition poses to these fields, with a special focus on its melodic and rhythmic aspects [37]. This work has implied the collaboration of musicologists, MIR researchers, developers and musicians in order to create a corpus based on the musical characteristics and the cultural context of the Arab-Andalusian musical tradition.

This corpus consists of three data collections:

1. audio recordings;

2. music scores;

3. lyrics.

These three collections are strictly linked. Audio recording was used as the main reference, since the music scores and lyrics are manual transcriptions of their predominant melody and sung lyrics. In the following sections, after a presentation of Arab-Andalusian music, the three collections are described according to each criterion. Furthermore, *ad hoc* interfaces developed to access and analyze the corpus' data and metadata are presented.

The author developed these interfaces during his mobility period at the Music Technology Group (MTG)[21] of the Universitat Pompeu Fabra in Barcelona, under the supervision of Barış Bozkurt and Xavier Serra (coordinator of the CompMusic project). Through these applications and the collaboration of Rafael Caro Repetto, Amin Chaachoo, as well as the supervisors, the corpus was thoroughly analyzed and improved (the complete description is provided in [37]). Furthermore, a first approach to the computational study of this musical tradition was developed, as presented in Section 2.8.

---

[21]`upf.edu/web/mtg` (Retrieved September 26, 2018

2.7.1   Arab-Andalusian Music

Arab-Andalusian music[22] is a music tradition formed around the 12th century in the Islamic territories of the medieval Iberian Peninsula known as Al-Andalus. With the migration of the Islamic Andalusian population to North Africa, this music tradition was disseminated in North African countries, where it has been preserved to date. The word "Arab" was added to distinguish this tradition from the contemporary music diffused in Spanish region of Andalusia. This tradition is essentially an indigenous Iberian music, which differs from Arab music traditions. For example, Arab-Andalusian music does not include microtones and *glissandi* which are typical of Arab music [41]. Nevertheless, some peculiarities have been inherited from Arabic poetry and music, such as language and poetic forms and the adoption of some melodic modes. Moreover, each country which indigenized Arab-Andalusian music as classic tradition developed some local characteristics. The main classical traditions are located in Morocco, Tunisia and Algeria. The following work focuses on the Moroccan tradition, known as *al-Āla*[23].

Arab-Andalusian music is performed through *nawabāt* (plural of *nawba*) which are suites of instrumental and vocal compositions ordered according to their metrical mode in a sequence of increasing tempo [37]. Originally, each *nawba* consisted of pieces composed in the same melodic mode named as *ṭāb‘* (plural *ṭubū‘*). A *ṭāb‘* is characterized by a diatonic scale built upon a fundamental note, within a certain range and including two or three predominant degrees and one or two sustained degrees, as well as a specific collection of characteristic melodic motives [42, 37]. Furthermore, each *ṭāb‘* is associated to specific emotional and spiritual content [41].

Moroccan Arab-Andalusian tradition fixes its repertoire in 26 *ṭubū‘*, but only 11 *nawabāt*. Some *ṭubū‘* are preserved only in fragmentary parts and their pieces are included in different *nawabāt*, according to the related *ṭubū‘* [76]. For this reason, each *nawba* has a primary *ṭāb‘*, and most of them have also secondary *ṭubū‘*. Nowadays, *al-Āla* tradition has also incorporated *ṭubū‘* of folk origin. One of them even forming a new *nawba*, as well as *makam* music from Eastern traditions [37].

A *nawba* can be also understood as a sequence of sung poems known as *ṣanā‘i‘* (plural of *ṣan‘a*). They are performed by mixed choir and orchestra in a heterophonic structure. The instrumental ensemble is generally formed by string instruments such as *‘ūd*, *rabāb*, *qānūn*, violin, viola, cello, double bass and piano, percussion instruments such as tar and darbuka, and occasionally also a clarinet as a wind instrument [37]. Either instrumental, both solo and orchestral, or vocal, both solo and choral are part of a *nawba*, preceding or interpolating the *ṣanā‘i‘*. These pieces are grouped in a performance according to each of the five rhythmic modes or *mawāzīn* (plural of *mīzān*) established in the tradition, and ordered according to the three possible renditions of the *mīzān* in increasing tempo [37]. The performance of a whole *nawba* can require several hours. Nowadays, only one *mīzān* of one *nawba* is usually performed. Each *mīzān* is performed in three tempi respectively named *muassa‘*, *mahzūz* and *inṣirāf*. These become structural sections and consequently are considered as forms.

---

[22]For a discussion about the terminology used to refer to this music tradition, please refer to [112]

[23]All the Arab terms are transcribed according to the standard proposed by [58], unless otherwise stated

2.7.2   Audio Recordings Collection

The audio recording collections consists of 164 long recordings, totalling more than 125 hours of music. Their length is variable, the average duration is around 45 minutes. The corpus described in the next sections represents the largest open source research corpus of Arab-Andalusian music tradition.

Coverage

The audio collection covers all the main instances of the four musical entities of Arab-Andalusian music:

- *nawba* (Table C.1);

- *ṭāb'* (Table C.2);

- form (Table C.3);

- *mīzān* (Table C.4).

As reported in Table C.1, all the eleven *nawabāt* of the Moroccan tradition are covered. Furthermore, the collection includes one *nawba* that does not belong to the classic tradition. This was recently created, as well as the other 21 recordings of the popular tradition that are not included in any of these *nawabāt*. The corpus covers 13 out of 26 *ṭubū'* of the *al-Āla* tradition, plus several other melodic modes derived from different origins. All the forms listed in Table C.3 belonging to the tradition with the exception of the *bugia*. All five traditional *mawāzīn* are covered, plus other rhythmic modes derived from different origins. Besides, 134 recordings contain sections that were annotated as not related to any *mīzān*.

Completeness

Each recording of the corpus is associated to a MusicBrainz[24] ID (MBID). An MBID uniquely identifies a recording, allowing to link its related data, such as score or lyric, and metadata. Furthermore, MusicBrainz framework offers the possibility to store all the metadata of each recording and then retrieve them using its API. A complete set of metadata is stored in MusicBrainz, from the recording title to the name of all the musicians that perform, including the related instruments. The metadata is stored in two versions: the original Arabic script and their transliterations, which are automatically generated on the basis of the American Library Association - Library of Congress (ALA-LC) standard for Arabic [131].

In order to facilitate the (computational) musicological analysis, each record was manually segmented by Amin Chaachoo, an expert musician and musicologist of this tradition. For each section it was annotated the *nawba*, *ṭāb'*, form and *mīzān*. Tables in Appendix C shows the number of sections annotated for each entity and the related overall duration. Only Table C.1 shows the number of recordings for each *nawba*. All the recordings contain pieces of a single

---

[24]`musicbrainz.org/` (Retrieved September 18, 2018)
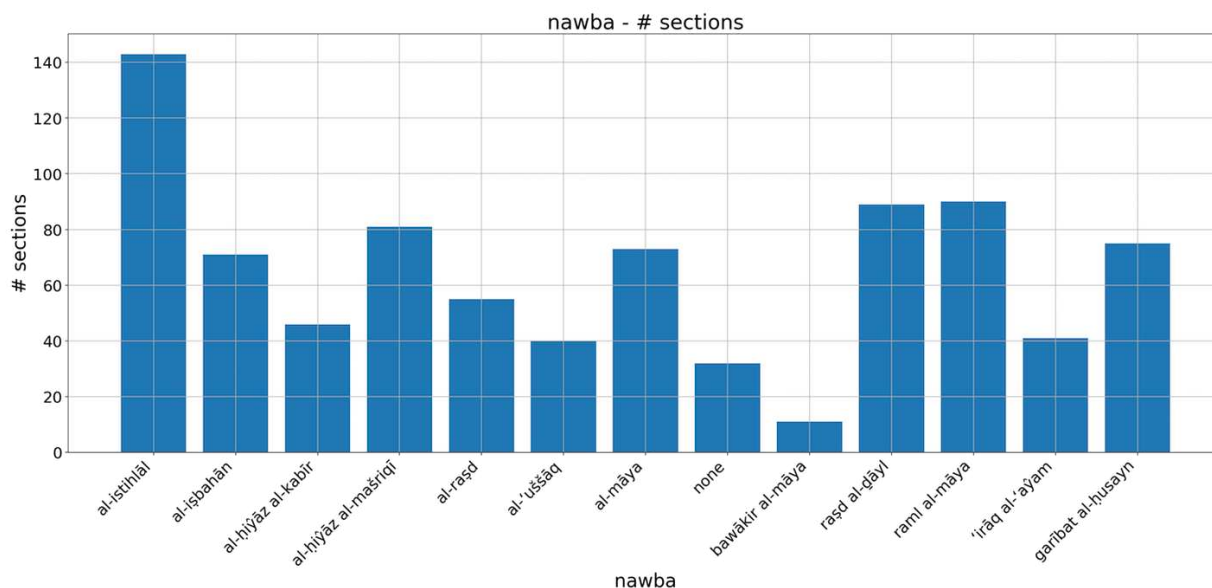
Figure 2.22: *Bar chart showing the number of sections per nawabāt as plotted in the Metadata Jupyter Notebook [37].*

nawba, with only one exception (with two *nawabāt*). This recording is not considered in statistics and in future analysis. However, Figure 2.22 shows the distribution of sections per *nawba*. Only six recordings in the corpus are part of a folk repertoire. These were not segmented and annotated and henceforth will be not included in the descriptions.

For each recording, several features were extracted, such as the fundamental frequency series and the pitch distribution, which are available in the recordings collection. The algorithm used to extract the first feature was originally developed for Turkish *makam* music [5], but the quality of the estimation was confirmed by a visual inspection of several samples by plotting the series together with the recording spectrograms. The recordings involve four categories of signals: vocal-only improvisation, single instrument improvisation and heterophonic multi-instrumental performance, with or without vocal. The pitch estimation quality is high in most parts of the audio with occasional octave errors mainly occurred during the low-pitch instrumental improvisation (e.g. 'ūd) [118]. The pitch distributions are extracted from the pitch series using a 7.5 cents resolution and smoothed using a Gaussian kernel with a standard deviation of 7.5 cents, as proposed in [7]. As result of the latter operation and the long duration of the recordings, the distribution is highly smooth.

Quality

The recordings collected in the corpus were chosen for their artistic quality and representativity in the Arab-Andalusian tradition. For copyright reasons and historical circumstances of this tradition in the last decades, most of them were recorded in the 1960s and 1970s. For this reason, the sound quality can be poor occasionally, with noisy and sporadically inaudible sections. Fur-

thermore, some recordings were made live events, therefore some disturbance from the audience, such as an applause, may overlap the music. In order to reduce the storage capacity and make the files easily downloadable, the recordings are stored in mp3 format, sampled at 44.1 or 48 kHz. There is no unique bit rate, however, the recordings usually have at least 128 kbps (with only six exceptions). The recordings can be either mono or stereo.

Re-Usability

As already mentioned, the entire collection is copyright-free and can be freely distributed. The whole collection is available in the Internet Archive[25], whereas the metadata are accessible in MusicBrainz[26]. In order to facilitate the browsing of the collection according to musically relevant concepts such as artist, *nawba*, *mīzān*, the online tool Dunya[27] [113] developed during the CompMusic project, allows to play the recordings and provides an API for retrieving the metadata and downloading all the data previously described. Furthermore, a set of Jupyter Notebooks (described in the Section 2.7.5) provides a more powerful and customizable tool for accessing all the data and metadata of the corpus. Finally, all the data can be directly downloaded as zip file from a Zenodo repository[28]. In this repository, each recording is stored in a folder named with its own MBID. Besides the mp3 file, each folder contains a JSON file with the editorial metadata from MusicBrainz, the transcribed score in XML format and two derived files, a txt file, importable into the spectrogram tool of Sonic Visualizer[29], with the fundamental frequency series extracted from the audio recording, and the pitch distribution in JSON format. The overall size of the zip file is approximately 9 GB.

2.7.3 Music Scores Collection

All the scores of this collection are manual transcription made by Amin Chaachoo. In doing these transcriptions, the focus is set on the stock of melodies orally transmitted that form the core of the al-Āla tradition. Even though divergences across performances are common, these are shared by all the orchestras. Therefore, improvisatory sections of instrumental and vocal solos have not been transcribed [37].

All the collection is available in machine readable music scores and linked with their corresponding recordings via the MBID. They were created using MuseScore[30], and then exported in MusicXML format.

---

[25]`archive.org/details/arabandalusian` (Retrieved September 18, 2018)
[26]`musicbrainz.org/collection/142ea0d7-7fdf-4ea5-9b04-219f68023d01` (Retrieved September 18, 2018)
[27]`dunya.compmusic.upf.edu/andalusian/` (Retrieved September 18, 2018)
[28]`doi.org/10.5281/zenodo.1291775` (Retrieved September 18, 2018)
[29]`sonicvisualiser.org/` (Retrieved September 18, 2018)
[30]`musescore.org` (Retrieved September 19, 2018)

Figure 2.23: *Excerpt of the music score for the recording with MBID e22549ae-4a0c-43ef-87f4-e0f81ed49d58, notating the performance of the mīzān qāʿim wa niṣf of nawba al-istihlāl [37].*

Coverage

Since the music scores consist in transcriptions of the recordings contained in the audio collection, the coverage is the same (see Section 2.7.2). The only exceptions are the sections corresponding to the forms *inšād*, *taqsīm* and *mawwāl*, since their melody is improvised, as well as the six not segmented recordings for which the scores are not available.

Completeness

All the metadata described in the audio recordings collection are linked to the scores by MBID. The lyrics are not included in the scores due to issues in deriving from the Arabic right-to-left writing system. However, the first line of each *ṣanʿa* is annotated inside the score with a textual label. The scores are not segmented in sections yet, but this long task is in process at the time of writing this work. Therefore, current annotations only include *nawba* and *mīzān*, which are unique per recording. Form and *ṭabʿ* will also be applicable once the segmentation process will be finished.

Quality

Arab-Andalusian music is performed in a heterophonic texture. Consequently, the scores contain the predominant melody underlaying the different performance layers (see Figure 2.23), disregarding occasional octave changes by specific vocalists and instrumentalists, and changes

between vocal and instrumental sections. They include tempo markings at the beginning and for each tempo change throughout the piece related to the performance in the transcribed audio recording.

Re-Usability

As for the audio recordings collection, music scores are completely open and accessible through different means. They can be viewed and downloaded from the MuseScore[31], a website which offers an online interface to play and interact with them. The description in MuseScore also contains the link to the corresponding recording in Internet Archive and MusicBrainz. Furthermore, the scores can be visualized in Dunya, which also provides the player to listen the related recording in the same windows. As for the recordings, the Jupyter Notebooks presented in the Section 2.7.5 allow to retrieve and download the scores according to more customizable research needs, as well as to visualize general statistics about the collection. Finally, they are included in the same Zenodo repository as the audio collection downloadable in zip file (see Re-Usability of Section 2.7.2).

2.7.4 Lyrics Collection

The *ṣanʿa* or sung poem is the fundamental form in a *nawba* performance and the main structuring element. Indeed, these poems are considered as one of the most valuable repertoires of Arab-Andalusian poetry. The relationship between lyrics, melody and emotional content associated with the corresponding *ṭābʿ* and *mīzān* confers them a high value for musicological research [37].

Coverage

The lyrics for the *ṣanāʿiʿ* of all the recordings of nine *nawabāt* in the corpus were identified. Currently, the lyrics collection contains machine readable lyrics for the audio recordings of only three *nawabāt*: *al-iṣbahān*, *al-māya* and *raml al-māya*. The remaining transcriptions are a work in progress.

Completeness

The collection contains the lyrics both in its original Arabic script and its transliterated version. As for the metadata, the transliterations are automatically computed according to the American Library Association - Library of Congress (ALA-LC) standard [131]. The lyrics are not aligned with the scores yet. However, they are linked with the corresponding section of the score in order to improve their usability and accelerate their analysis. For the same reason, each *ṣanʿa* was divided into lines and line sections, and stored both in TSV and JSON formats.

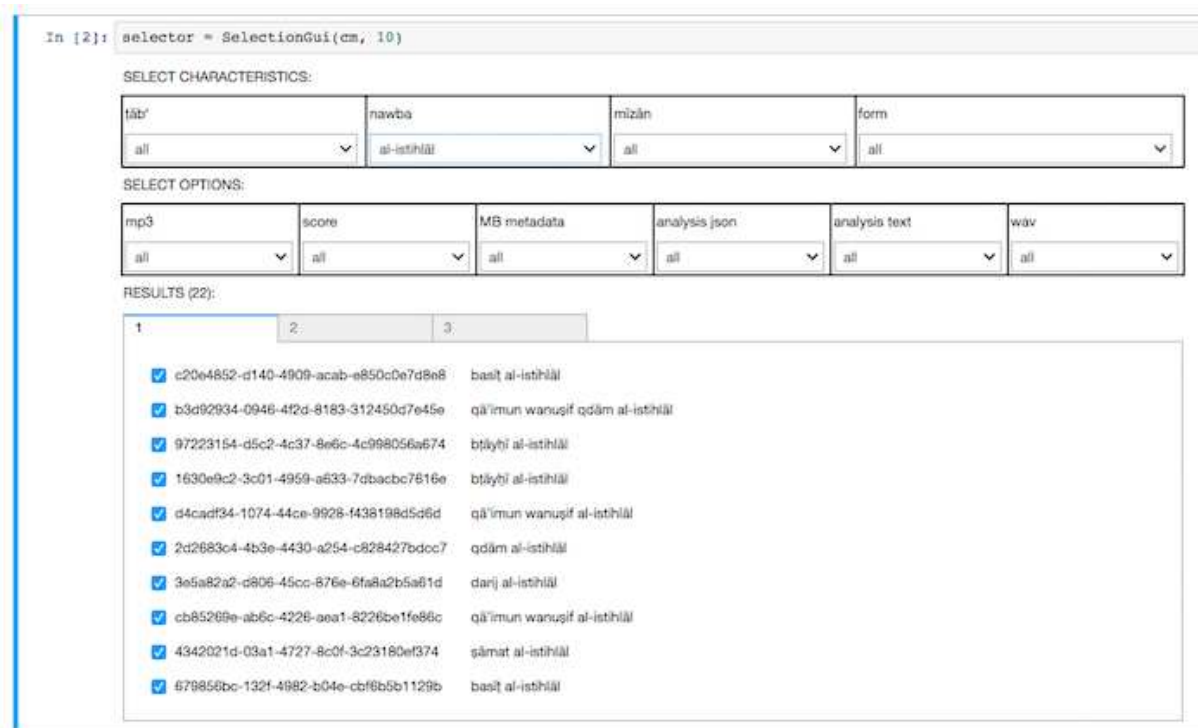[31]`musescore.com/mtg` (Retrieved September 19, 2018)

Figure 2.24: *Selection tool of the Corpus Jupyter notebook.*

Quality

The lyrics sung in a particular performance might change due to several reasons, such as the school to which the orchestra belongs to, the social context of the performance or the personal criterion of the director of the orchestra. The lyrics contained in this collection have been manually typed by Amin Chaachoo, who listened to each recording and selected the corresponding *ṣanʿa* as compiled by the musician and scholar Mehdi Chaachoo, whose compendium takes into account the current differences across schools within the al-Āla tradition [37].

Re-Usability

The lyrics, both in their original Arabic script and their transliterations, are available both in TSV and JSON formats. The whole collection is freely downloadable from Zenodo[32].

2.7.5   Accessing to the corpus

In order to easily access and exploit the Arab-Andalusian corpus, a set of Jupyter notebooks[33], written in Python, was developed and shared in a public GitHub repository[34]. These notebooks

---

[32] doi.org/10.5281/zenodo.1291903 (Retrieved September 19, 2018)

[33] jupyter.org/ (Retrieved September 19, 2018)

[34] github.com/MTG/andalusian-corpus-notebooks (Retrieved September 19, 2018)

provide a simple way for accessing to data and metadata of the corpus and a set of visual tools composed by widgets[35] to visualize data and metadata as well as to perform repetitive tasks in a safe way.

This repository contains four notebooks [37]:

- **Corpus**: to select, download and compute the corpus data and metadata;

- **Metadata**: to group, visualize and analyze metadata;

- **NawbaPitchAnalysis**: to visualize pitch distribution and note/class distribution of a single recording or of a group of them;

- **NawbaRecognition**: to compute several experiments to evaluate the performance on *nawba* recognition of algorithms based on templates derived from scores.

A common section at the beginning of each notebook allows to download all the metadata from Dunya. In order to fully exploit the notebooks' tools, this section also create a Python object to easily manage them. A class offers several methods which can help developers that firstly approach this corpus.

The first notebook provides three main functionalities:

1. a widget to select a list of recordings, using filters and checkboxes;

2. a widget to download mp3, scores in XML and metadata from MusicBrainz (in JSON format);

3. a widget to compute the pitch-related characteristics, namely the fundamental frequency series (unfiltered or filtered with the algorithm proposed in [5]), pitch distribution and several estimations of the tonic frequency.

Each visual tool can also be used by musicologist with no experience in coding.

Figure 2.24 shows the first tool. Several filters can be applied through drop-down lists. The first line provides the possibility to filter by the four musical entities previously described: *nawba*, *ṭābʿ*, form and *mīzān*. The second line of filters allows to verify if the mp3s, scores, metadata and computed files are stored in the computer. Furthermore, the user can select the desired recordings through single checkboxes. The result of this selection can be easily stored in a Python list containing all the MBID of the selected recordings.

This list is the input of both the second and third widgets. The first one consists of a simple interface through which the user can select the kind of files to download (Figure 2.25a). The three options are: mp3, score in XML and the metadata from MusicBrainz. With a simple click, the selected file categories can be downloaded for the list of recordings passed by input. An additional function allows to overwrite files even if they are already stored in the directory.

The last widget is presented in Figure 2.25b. Similarly to the previous widget, this provides the possibility to analyze the recordings. The first result of this analysis is the pitch profile

---

[35]`ipywidgets.readthedocs.io/en/latest/` (Retrieved September 19, 2018)

(a)

**(b)**

Figure 2.25: *Two widgets to (a) download data and metadata, and (b) compute pitch analysis.*

of the selected recordings, extracted using an algorithm originally developed to analyze Turkish *makam* music [5]. As mentioned in Section 2.7.2, from this fundamental frequency series the pitch distribution is computed. The resulting pitch distribution is consistently smooth, as shown in Figure 2.26. Furthermore, for each selected recording the tonic frequency is estimated
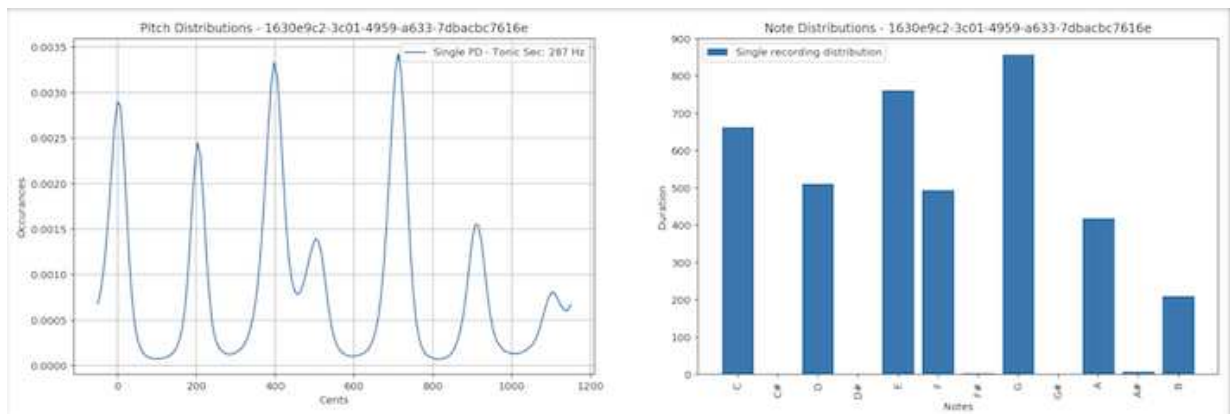
Figure 2.26: *Visualization of the pitch distribution computed from audio and of the pitch class distribution from the scores of a single recording as plotted in the NawbaPitchAnalysis Notebook.*

through three algorithms. Two of them use the algorithm proposed in [6] to the filtered and unfiltered version of the fundamental frequency series, whereas the last one applies the algorithm to each section of the recordings and evaluates the most frequent value. All the elaborated data is stored in JSON format, but for the fundamental frequency series it is possible to create a text file with all the values importable in the spectrogram tool of Sonic Visualizer. By selecting the

```
In [2]:  cr = CrossMetadataVisualization(cm)
```

Column: `mizan ▾`   Row: `nawba ▾`   Title: `transliterated_name ▾`   Statistics: `# sections ▾`

Results:

mizan / nawba	none	basīṭ	bṭāyḥī	ternary	binary	binary and bṭāyḥī	ḥaddārī	darŷ	zandānī and quddām	qā'im wa niṣf quddām	
al-istihlāl	44	8	26	0	15	0	0	9	0	27	14
al-iṣbahān	26	5	12	0	1	0	0	0	0	15	12
al-ḥiŷāz al-kabīr	17	7	9	0	2	0	0	1	0	5	5
al-ḥiŷāz al-mašriqī	32	6	11	2	15	0	0	3	0	0	12
al-raṣd	19	5	9	0	0	0	0	0	0	0	22
al-'uššāq	13	5	2	0	2	0	0	0	0	9	9
al-māya	24	19	10	0	2	0	0	2	0	3	13
none	4	0	1	0	1	0	1	0	5	0	20
bawākir al-māya	2	0	1	0	0	1	0	0	0	0	8
raṣd al-ḍāyl	31	19	8	0	1	0	0	0	0	14	15
raml al-māya	31	15	15	0	0	0	0	0	0	12	17
'irāq al-'aŷam	11	2	7	0	4	0	0	2	0	0	17
garībat al-ḥusayn	26	14	4	0	7	0	0	0	0	12	10

Figure 2.27: *Interactive table showing the number of sections per nawba and mīzān as plotted in the Metadata Jupyter Notebook.*

option *Pitch analysis in txt*, the values of the pitch profile are exported in a txt file, importable in the spectrogram tool of Sonic Visualizer. As explained in Section 2.7.2, a txt file with the filtered fundamental frequency series and a JSON file with the pitch distribution are downloadable from Zenodo (see Section 2.7.2), whereas the other derived files are only computable with the notebook.

The second notebook has got three other widgets. The first one allows to group the recordings or their sections by the main musical entities of Arab-Andalusian music (*nawba*, *ṭāb'*, form and *mīzān*) and to show the main statistics concerning the number of sections, recordings and the overall time. This information can be also visualized in bar charts, such as the one reported in Figure 2.22. This tool is fundamental to have an overall view of the corpus. The second widget provides a tool to cross information for two different characteristics and visualize the results in an interactive table (see Figure 2.27). By selecting the musical entities in row and column, the information is crossed and the number of sections, the overall or average time of each pair of characteristics is visualized. The last one provides a simple function to visualize what *nawba*, *ṭāb'*, form and *mīzān* belongs to each section of a selected recording.

The third notebook provides a simple tool to visualize both pitch distributions computed from audio recordings and pitch class distributions from the scores, such as in Figure 2.26. This widget offers the possibility to compare the distributions of a single recording with the average distribution of a single *nawba*, centering the distribution in three octaves as well as folding the distributions in a single octave.

The last notebook is dedicated to a *nawabāt* recognition experiment which is throughly described in the next section. This experiment provides the possibility to replicate the experiment in its entirety.

2.8 Nawba recognition

The first experiment that exploit the potential of the Arab-Andalusian music research corpus is related to the the automatic recognition of *nawabāt*, whose methodology and results are published in [118]. The approach relies on template matching applied to pitch distributions computed from audio recordings to match with templates learned from a score collection. Template matching is widely used in the computer vision field [26]. and in various content-based music retrieval tasks since early days of MIR [71, 140]. Following the influential study on tonality perception of [88], distribution template matching became largely used for automatic tonality or modality detection. The most commonly used features for matching are note distributions or pitch class distributions [107, 44, 138].

The algorithm presented in this work is the first computational approach in literature that uses distribution template matching for *nawba* detection in Arab-Andalusian music context. In order to carry out the study, a dataset from the overall corpus was extracted. The next sections describe the methodology, dataset, experiment, and results concerning the use of a first algorithm for the automatic *nawba* recognition task.

2.8.1 Methodology

The novel approach to *nawba* recognition proposed in this section uses templates computed from music scores. The core idea is to compare the pitch distribution of an Arab-Andalusian recording with several templates in order to discover the *nawba* to which this recording belongs. The input data of this methodology is given by several scores for every *nawba* in order to build the templates and audio recordings for which the *nawba* is unknown. From each score, a pitch class distribution in total duration is computed through the algorithms described in the previous section. These distributions are further folded in a interval of twelve semitones (one octave). The resulting twelve bar distributions of the recordings of a same *nawba* are averaged and normalized to a total sum of 1. The templates are synthesized from the pitch class distribution using a Gaussian curve for each distribution value. To obtain a normalized distribution comparable to the pitch distribution of a single recording, the values of a bar p are considered as the area of each Gaussian curve $g(x)$ The following formula is used in order to calculate the area under the curve:

$$p = \int_{-\infty}^{\infty} g(x)dx = \int_{-\infty}^{\infty} ae^{\frac{-(x-b)^2}{2c^2}} dx$$

,
where a is obtained as follows:

$$a = p\sqrt{2\pi c^2}$$

,
where c is the standard deviation (in the experiments considered in a range between 20 and 40 cents) [118]. The variable b is the average value and center of each curve. Each curve is positioned with intervals of 100 cents. An example of the resulting template can be observed in
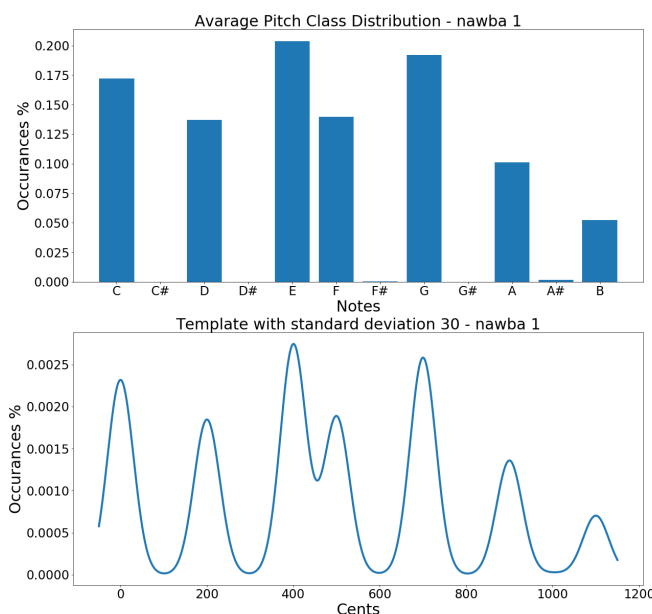
Figure 2.28: *An example of average folded pitch class distribution for* nawba al-istihlāl *and the corresponding template synthesized by using a standard deviation value of 30 cents.*

Figure 2.28. The template is normalized to 1. A template is created for every *nawba*. Each of them are compared with the pitch distribution of a single track.

The pitch distribution is extracted from the pitch profile as described in Section 2.7.2. The pitch distributions of the recordings are computed using a 7.5 cents resolution and smoothed using a Gaussian kernel with standard deviation of 7.5 cents. The smoothness of this distribution can be seen in Figure 2.29. In this plot, it can also be observed that the pitch distribution is folded in one octave as for the templates. This folding operation requires a reference value (an origin). Usually, the tonic frequency is the most commonly used value, but in this case is unknown. To compensate for this lack, the tonic frequency was substituted by the frequency of the maximum peak of the distribution. Nevertheless, the choice of this value does not affect the algorithm, since the template matching involves rolling the pitch distribution completely in one octave. In order to find the template with the best match, the pitch distribution is shifted and the distance is calculated at each shift. The minimum distance refers to the best match. Figure 2.29 shows an example of good matching between a *nawba* template and the pitch distribution of a recording.

2.8.2 Dataset and experiment

In order to assess the methodology delineated in the previous section, a dataset of 77 long recordings, totaling more than 58 hours of music, was extracted from the overall corpus. The complete
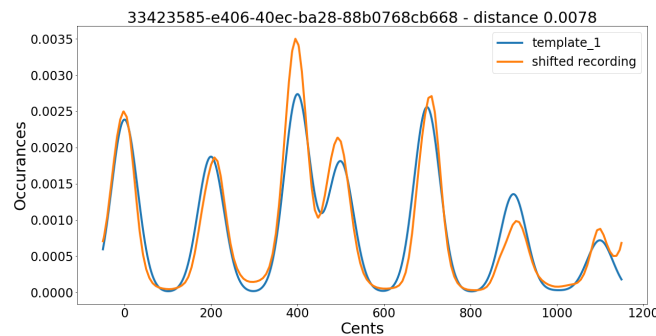
Figure 2.29: *An example of comparison between a recording pitch distribution and the* nawba al-istihlāl *template.*

Table 2.22: *Overall duration of the recordings for each nawba in the nawabāt detection dataset.*

Nawba name	Duration
al-istihlāl	5h 41'
al-iṣbahān	6h 20'
al-ḥiŷāz al-kabīr	5h 22'
al-ḥiŷāz al-mašriqī	3h 37'
al-raṣd	5h 18'
al-ʿuššāq	4h 35'
al-māya	6h 01'
raṣd al-ḏāyl	5h 49'
raml al-māya	6h 23'
ʿirāq al-ʿaŷam	4h 13'
garībat al-ḥusayn	5h 26'

set of data and metadata of the corpus is essential to obtain a reliable ground-truth for the experiment. Those recordings that do not belong to a canonical *nawba* were excluded and the remaining ones were selected according to two criteria: audio quality and length of the recording [37]. The only recording with two *nawabāt* in the same audio file was excluded *a priori*. The audio quality criterion involved an auditory evaluation of several samples of each recordings. The second criterion — the length of the recording — is derived from music theory, in the sense that, if the performance of the *mīzān* of a *nawba* requires almost one hour, a track of few minutes cannot adequately represent such a performance. Since *nawabāt al-ḥiŷāz al-kabīr*, *al-ʿuššāq* and *ʿirāq al-ʿaŷam* only have seven recordings each, and considering the necessity of a weighted dataset, these cases imposed the maximum number of recordings for each *nawba*. Therefore, the resulting dataset includes seven recordings for each of the eleven *nawabāt*. The overall duration of the recordings for each *nawba* is presented in Table 2.22.

The experiment divided the dataset in two stratified random subsets, composed as follows.
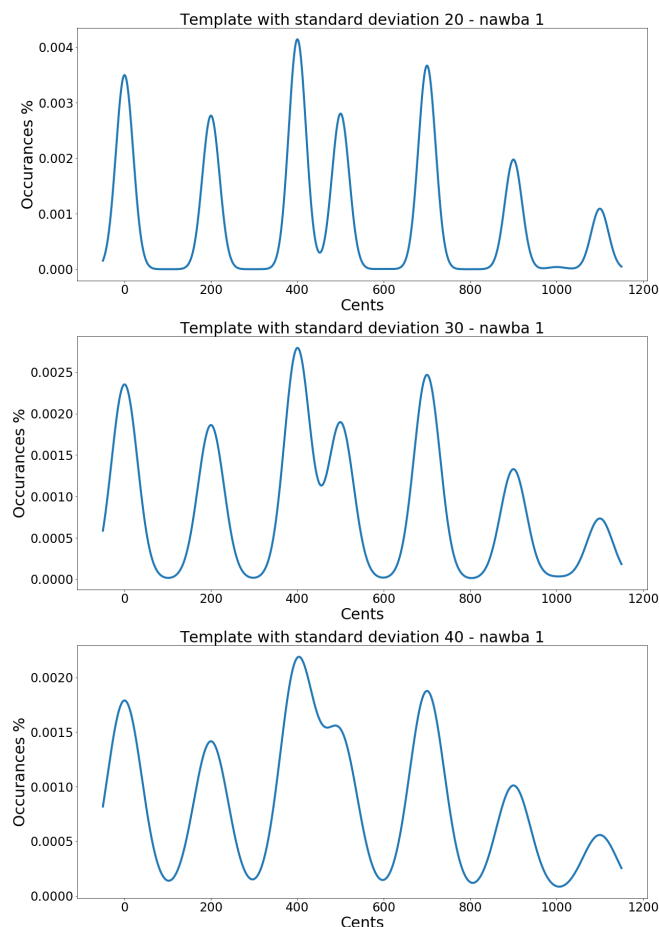
Figure 2.30: *Three templates for* nawba al-istihlāl *synthetized using respectively 20, 30 and 40 cents as standard deviation value.*

For each *nawba*, the scores of six recordings train the templates, whereas the remaining one is part of the test set. This procedure was critical to make the template independent from the test recordings. The experiment was repeated seven times: each time a different recording for every *nawba* was selected. Therefore, each recording was tested once.

The standard deviation value characterizes the width of the "bell" of the Gaussians functions and strongly affects the performance of the recognition. The tested values of standard deviation are three: 20, 30 and 40 cents. Figure 2.30 clarifies the great impact of this value on the templates shape. Another factor that highly affects the performance is the distance metric used to match the pitch distribution with the templates. The distance metrics tested in the experiment are: City-Block (L1), Euclidean (L2), the inverse of the correlation and Canberra.

As mentioned in Section 2.7.5, with a Jupyter notebook whose code is openly shared, the experiment can be replicated.

Distance	Standard Deviation		
measure	20 cents	30 cents	40 cents
City-Block (L1)	0.727	0.740	0.688
Euclidean (L2)	0.740	0.753	0.688
Correlation	0.377	0.390	0.390
Canberra	0.377	0.701	0.571

**Table 2.23:** Nawba *recognition performance of the proposed algorithm.*

2.8.3 Results

The overall results are provided in Table 2.23 and can be considered as a good starting point for this task with all the eleven *nawabāt* of the classic tradition. The Euclidean (L2) distance measures and templates synthesized using a standard deviation equal to 30 cents provided the best performance. In general, this distance metric resulted in a higher performance than the others. Only through a 40 cents standard deviation, the Euclidean (L2) and City-Block (L1) distance have equal results. Considering the overall performance, the standard deviation of 20 and 30 cents gave best results. Moreover, Euclidean (L2) distance resulted be more suitable for the *nawba* recognition task. However, the overall behavior of City-Block (L1) is similar to the Euclidean (L2) one. The correlation led to the lowest average performance, and did not seem suitable for this kind of analysis. The performance observed for the Canberra distance was highly affected by the standard deviation through which the templates were developed, leading to good results only when the standard deviation was set at 30 cents.

Figure 2.31 shows the overall confusion matrix obtained by summing all the results of the seven folds for the best combination of standard deviation (30 cents) and distance metric (L2). Considering the overall result of 75% of recognized *nawabāt*, the distribution of the majority of values in the diagonal was expected. In general, the values outside the diagonal did not seem to follow a precise pattern. The worst results are obtained for *al-'uššāq* (with Dunya id 6[36]), that is the only *nawba* with performance lower than 50%. Figure 2.32 shows an example of mismatch for a recording of this *nawba*. All the results and the plots concerning the best experiment are available in the GitHub repository[37].

---

[36]This id is related to the first version of the metadata, now deprecated and substituted with an unique id. The id is maintained in this work to facilitate the comprehension of the results

[37]github.com/MTG/andalusian-corpus-notebooks (Retrieved September 19, 2018)
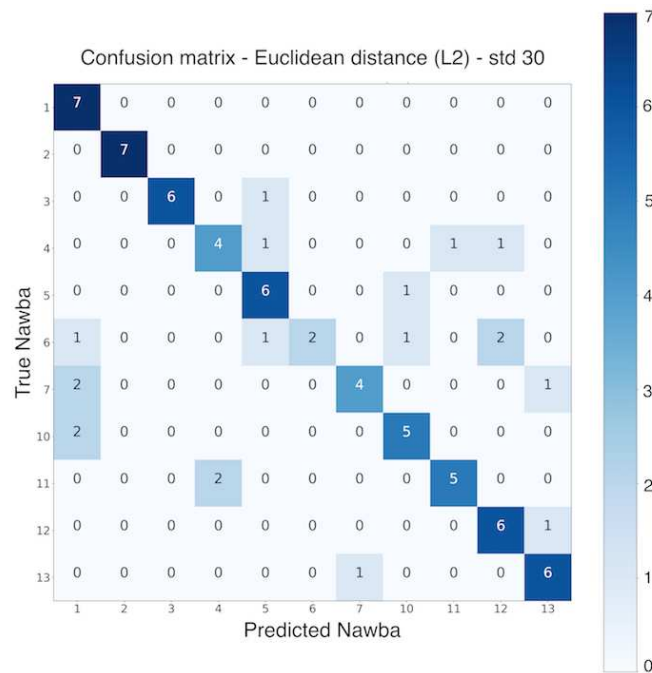
Figure 2.31: *Overall confusion matrix for the experiments with Euclidean distance and 30 cents as standard deviation value.*
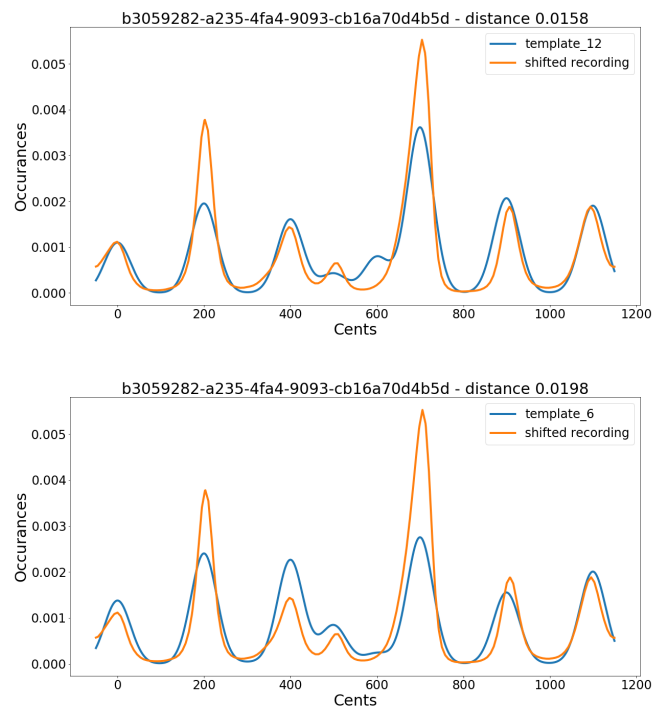
Figure 2.32: *An example of incorrect recognition of the* nawba. *The track is recognized as belonging to* nawba *12, but the correct* nawba *is 6.*

Chapter 3

Historical Musical Instruments

3.1 Case study: Archaeology and virtual acoustics

As introduced in Section 1.2.2, the musical heritage includes scores, audio recordings as well as musical instruments. There are several kinds of musical instruments, each with peculiarities that have to be preserved and communicated. However, all of them share a common characteristic: they have to be played to be understood.

The aim of the project "Archaeology and virtual acoustics"[1] is to valorize an exceptional archaeological find recovered during the recent reassessment of the MSA in Padova. The artifact is an ancient Pan flute, probably of Greek origins, consisting of 14 reeds of different lengths held together by ropes and a natural binder, and originally coated with a resin layer (now partially missing) [8].

The artifact arrived in Padova thanks to Carlo Anti, who directed the Italian Archaeological Mission in Egypt since 1928. Carlo Anti held the position of Rector at the University of Padova, and contributed to renovate and modernize the university and its buildings. Among them, Palazzo Liviano, which is the current location of the MSA. The archaeologist also led excavations in the ancient village of Tebtynis in the Fayum oasis, from 1930 to 1936.

At the moment, the origin of the Pan flute is not completely certain. This ancient musical instrument was stored in a box, originally made for photographic plates (see Figure 3.1), belonged to Gilbert Bagnani, Italian-English archaeologist who assisted Anti. The box cover reports a French sentence in the tiny handwriting of Bagnani's wife, which sets the original finding in Saqqara, in the area of the Mastaba n. XV, thus near Pepi II's tomb [8].

Further information is found in Anti's archive and in a letter written by Evaristo Breccia (Director of the Archaeological, Museum of Alexandria), where he asked to Anti about this instrument which he saw during a visit in Tebtynis.

This hypothesis of origin is supported by the presence in Padova of other antiquities from Bagnani's campaigns, stored in small boxes similar to the Pan flute one, and unlike other archaeological materials. Except for a few exceptions, the findings were recovered at the MSA in 1935, therefore this probably corresponds to the year of discovery of the Pan flute [8].

---

[1] financed by the University of Padova, under grant no. CPDA133925

Figure 3.1: *The Pan flute in the box for photographic plates, before restoration (photo by Team EgittoVeneto).*

(a)   (b)

Figure 3.2: *The Pan flute after the restoration (photo by Nicola Restauri).*

After a major restoration program for consolidation and preservation, the Pan flute was first exhibited in April 2013, during the exhibition "Egypt in Veneto", hosted at the MSA [100]. Figure 3.2 shows the musical instruments right after this delicate operation. The restoration program also allowed to obtain the first analytical investigations, such as infrared (IR), ultraviolet (UV) and X-ray analysis. These were needed to test the preservation status as well as to test the construction techniques and chemical components. Any evidence of earlier decorations is not found. The archaeological find is currently exhibited at MSA, in a dedicated show-case with air-tight and continuous monitoring of the environmental conditions [8].

In the next sections, the analysis performed and the resulting information extracted from the measurements are analyzed.

This multidisciplinary project, coordinated by Paola Zanovello of the Department of Cultural Heritage (DBC), involved a huge group of researchers. At DBC, the coordinator in collaboration with Giulia Deotto and Alessandra Menegazzi (curator of MSA) are responsible for the archaeological study of this ancient artifact. The analysis presented in the next section were performed by a team composed by Ivana Angelini, Cinzia Bettineschi, Emanuela Faresin, Gianmario Molin and Giuseppe Salemi. The team of the CSC is composed by Federico Avanzini, Sergio Canazza, Giovanni De Poli, Carlo Fantozzi, Edoardo Micheloni, Antonio Rodà and the author. The main role of the author was to develop the multimedia installation presented in Section 3.2.2. This installation concept was created with the collaboration of the designer Silvia Gasparotto of the University of San Marino, with which the author defined the methodology presented in Section 3.2.1. Furthermore, the author was actively involved in the measurements phase and the study of the tuning, as well as in the definition of the questionnaire for the assessment, presented respectively in Sections 3.1.2, 3.1.3 and 3.2.3.

3.1.1 Analysis

As mentioned in the previous section, several non-invasive analysis have been performed. A complete description of the non-destructive mineralogical investigations of the pipe coating through X-ray Diffraction (XRD) coupled with Scanning Electron Microscopy (SEM) and Energy Dispersive Spectroscopy (EDS) can be found in [8], with several other analysis. The last analysis carried out consists in the radiocarbon ^{14}C, which dated the archaeological find at 7 century a.C.

A 3D model of the Pan flute was acquired through a non-invasive scanning (without any contact) of the surface and borders. The measurement device was a ScanArm V3 from Faro, a seven-axis measurement system that has a fully integrated laser scanner with a scan rate up to 19200 points/s, an accuracy of ± 35 μm. In [8], the scanning procedure can be examined in depth. The resulting high resolution model counts 920152 triangles and it is shown in Figure 3.3.

A first set of measures were gathered, in order to have a first estimation of the overall dimensions. The musical instrument has closed pipes, therefore an internal analysis was required in order to have precise measures of each pipe. Furthermore, it is known that the internal lengths of the pipes are reduced by carefully increasing the thickness of the closed ends through the addition of wax or propolis, in order to fine tune fundamental frequencies [46]. Despite the restoration of the ancient Pan flute, some pipes are still partially obstructed. Therefore, the interior of these
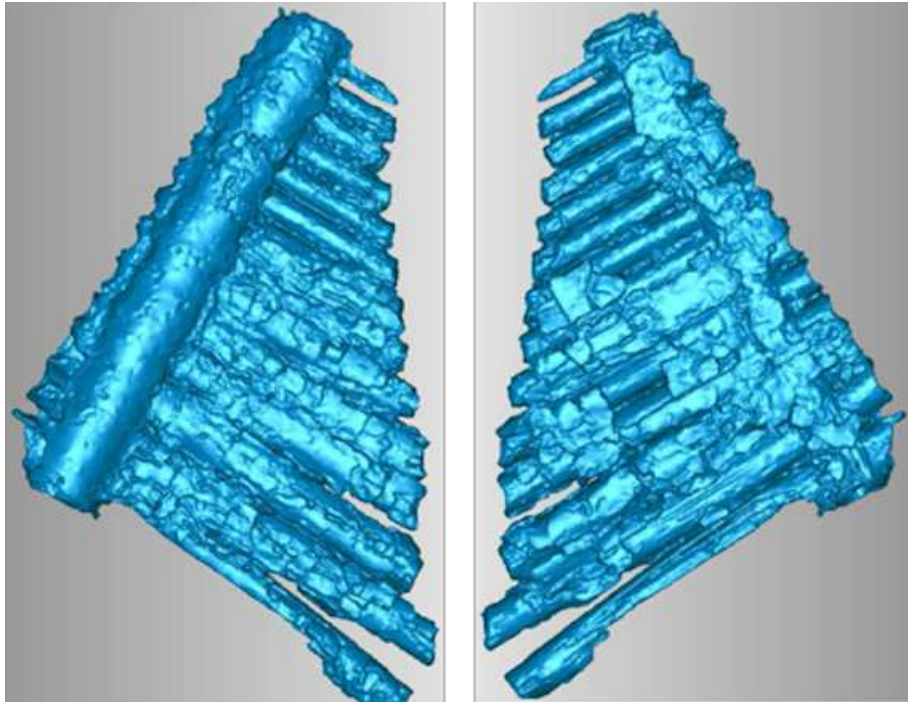
Figure 3.3: *The 3D model of the ancient Pan flute.*

pipes is not completely visible and not directly inspectable [9]. In order to refine the measurements, a computerized tomography (CT) scan was performed with a GE LightSpeed VCT 64 Slice CT scan[2]. The next section illustrates how the measurements were gathered and how the scanning was used to estimate the intonation of the Pan flute.
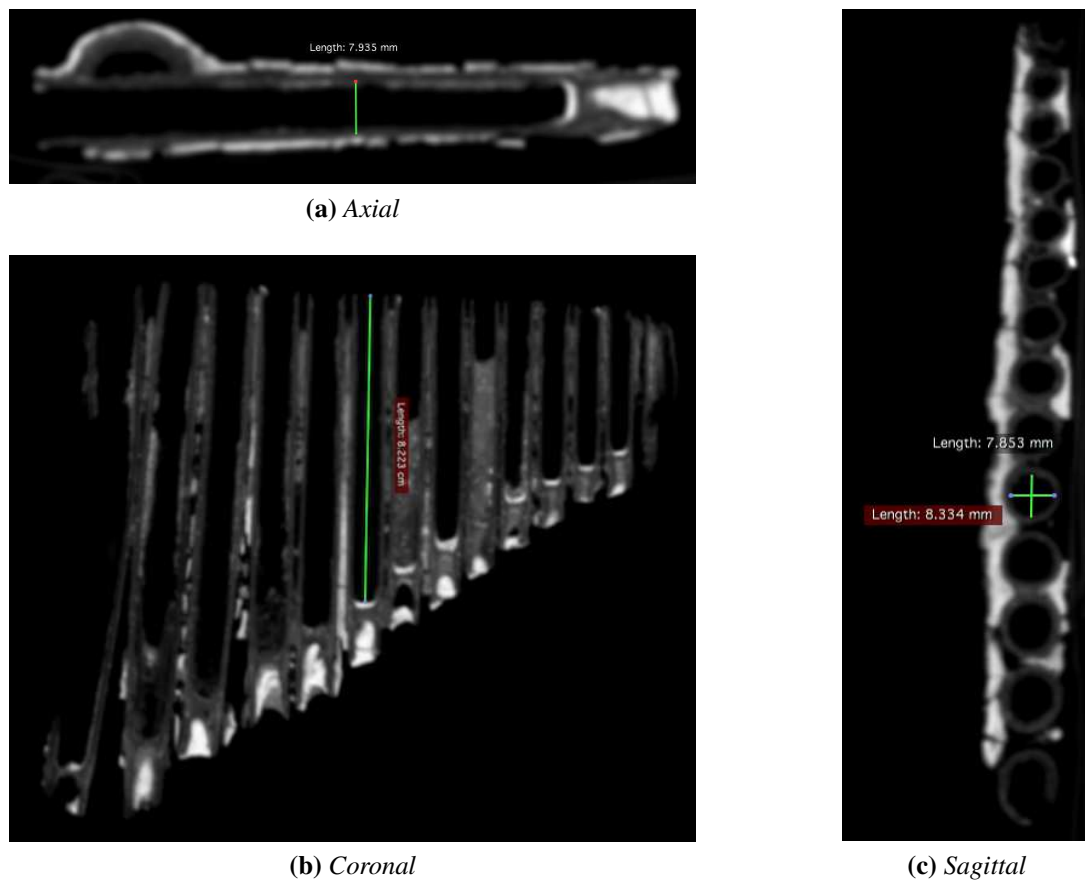
3.1.2   Measurements

The CT scanning was read through the open-source software Horos. This medical image viewer also provides tools to extract reliable measures from the scan. In order to browse inside the three-dimensional image and to perform precise measurements, two main views were used: a 3D MultiPlanar Reconstruction (MPR) and a 2D orthogonal MPR. The latter view has three orthogonal planes, defined as axial, coronal and sagittal. Figure 3.4 shows an example of the three views related the 3D scan of the Pan flute.

The main objective of the study was to determine the fundamental frequencies of each pipe. The required measures are internal length and internal diameter of the pipe. Although the exceptional preservation status of the Pan flute, some parts are damaged or corrupted. With the aim to reduce the errors induced by this alterations, a total of eighteen measures were collected for each pipe.
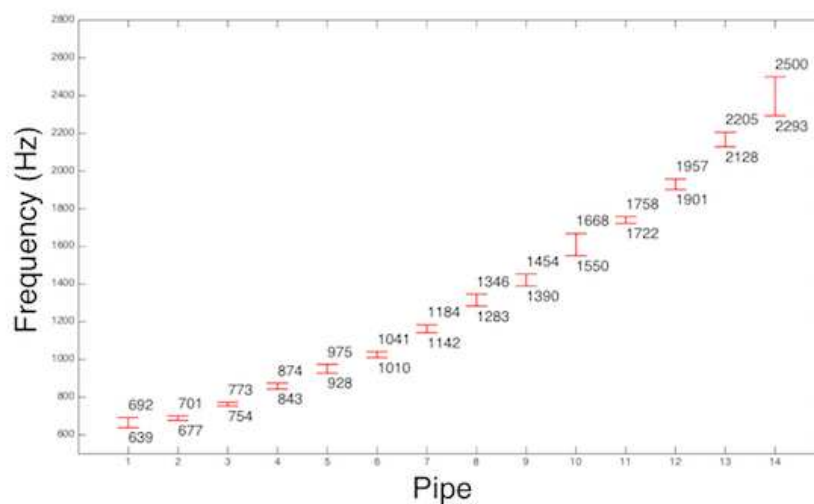
In order to find the internal length, a total of six measures were collected for each pipes. The

---

[2]dutiful thanks are due to the radiologist Carlo Rettore and its team, who performed the CT scan of the Pan flute at the Hospital of Cittadella (PD)

(a) *Axial*

(b) *Coronal*

(c) *Sagittal*

Figure 3.4: *Example of measurements with the three views: axial 3.4a, coronal 3.4b and sagittal 3.4c.*

Figure 3.5: *Fundamental frequencies (min and max) estimated for each pipe starting from the measurements taken from the CT scan.*

pipe openings are not straight, but they have a slightly u-shaped at one side in order to provide an embouchure to the player. It is possible to notice this peculiarity by carefully observing the embouchure of the musical instrument in Figure 3.2b. In the axial view, it is possible to measure the pipe using minimum and maximum points of the opening. Due to irregular shape of the pipes, two measures were gathered for each starting point: the minimum and maximum points at the bottom of each pipe, for a total of four measures. Through the same methodology, two other measurements were performed in the coronal view. In this plane, the embouchure is not visible, therefore only one starting point is used. Figure 3.4b shows an example of measurement in this plan.

In order to measure the pipe diameters, twelve measures were gathered. The problem tackled in this set of measurements is the irregularity of the pipes. In some areas, their sections are oval-shaped rather than circular. For this reason, a total of twelve measures were collected. For each plane, three points were selected: the top, middle and bottom of each pipe. One measurement for each point was performed both in axial and coronal planes. For the first panel, one example of measurement in the middle of the pipe can be observed in Figure 3.4a. Two measures were taken sagittal plane, as can be seen in Figure 3.4c.

The measurement process highlighted several issues that required some subjective interpretations. For example, the longest pipe is curved and evidently broken. The problem was resolved by using the 3D Curved-MPR tool, which provided a set of reference points on the curve, virtually straightening the pipe [9]. In some cases, it was difficult to discern the edge of internal surface from obstructing material. Furthermore, it was not possible to perform a direct measure from the openings of the shortest pipe due to heavy damages. Consequently, the opening positions were estimated from the neighboring pipes. An error that mostly impacts the measures was the CT scan resolution: every voxel (volumetric pixel) is isometric and it measures 0.625 mm [9]. However, the redundancy of the measures provided a plausible estimation of the lengths and diameters of each pipe.

3.1.3   Tuning

The internal length and the diameter allow to estimate the fundamental frequency of each pipe, according to the following equation:

$$f = \frac{c}{4(l + \Delta l)}(Hz), \qquad (3.1)$$

where c is the sound velocity, l is the internal pipe length, and $\Delta l \sim 0.305d$ is the length correction at the open end, proportional to the internal pipe diameter d [70].

In order to consider the measurements issues described in the previous section, the minimum and the maximum values for the internal length and diameter were used. The result is a range of frequencies for each pipe. Figure 3.5 shows the estimation of the fundamental frequencies. The minimum value of frequency was calculated from the previous formula by using the maximum values of length and diameter. Vice versa, the maximum values were calculated from the corresponding minimum values. From those frequency ranges some interesting patterns can be identified, as explained to follow.
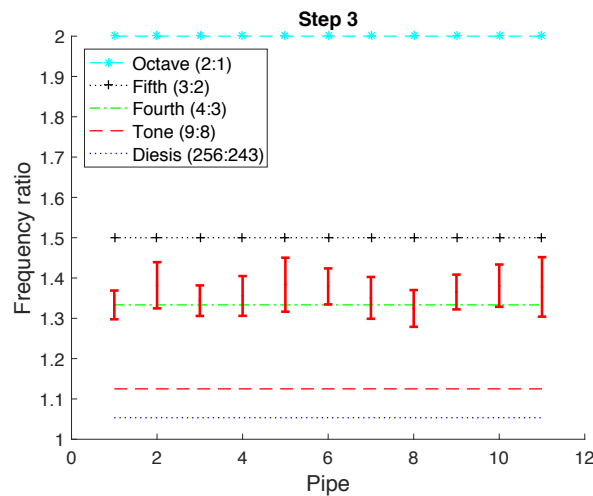
**Figure 3.6:** *Pitch ratios calculated as $f(n+3)/f(n)$, where $f(n)$ is the fundamental frequency of the nth pipe. The horizontal lines correspond to the basic theoretic intervals.*

According to musicologists [77], the ancient Greek music system was based on the *tetrachord*. A tetrachord is a group of four notes where the ratio between the pitches of the fourth and the first note is equal to 4:3, namely a perfect fourth. Figure 3.6 shows the pitch ratios calculated as $f(n+3)/f(n)$ for $n = 1, 2, ..., 11$, where $f(n)$ is the fundamental frequency of the nth pipe. For each couple of pipes a range of values was computed in order to manage the errors propagation. The horizontal line represents the 4:3 ratio (dot-dashed line) By comparing the lines, it is possible to observe how the intervals are compatible with the tetrachord definition. Three *genera* of tetrachord exist, based on their pitch intervals: diatonic, chromatic and enharmonic. For example, the diatonic tetrachord is characterized by intervals that are less than or equal to half the total interval of the tetrachord. Usually, this tetrachord begins with one small interval followed by two bigger intervals, more or less corresponding to a tone (9:8). Figure 3.7 shows the pitch ratios among adjacent pipes, i.e $f(n+1)/f(n)$. Some compatible intervals with a tone (9:8) are recognizable. Smaller ones are compatible with what some musicologists call *diesis*, corresponding to the ratio 256:234. Two tetrachords can be joined by following two different schemas, called *synaphē* (conjunction). This happens when the top note of the lower tetrachord corresponds to the bottom note of the higher one, and *diazeuxis* disjunction, when there is an interval of a tone between the tetrachords. By observing the sequence of intervals of Figure 3.7, some joint tetrachords can be recognized: e.g., the pitch of the first eight pipes are compatible with two disjunct tetrachords, as represented in Fig. 3.8 [9].

In spite this musicological observation, the fundamental frequencies employed to reconstruct the sound of the Pan flute in the multimedia installation described in the next section are average values of the estimated frequencies. This choice aims to avoid the risk of creating a fabrication of history. Nevertheless, other musicological studies will be carried out in order to find a reliable estimation of the ancient musical instruments.
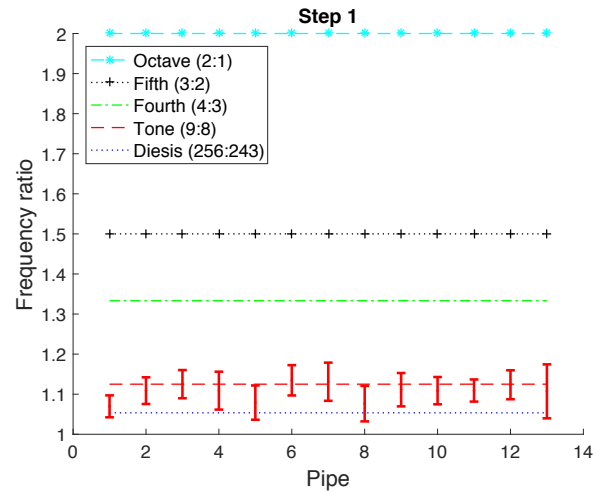
Figure 3.7: *Pitch ratios calculated as $f(n+1)/f(n)$, where $f(n)$ is the fundamental frequency of the nth pipe. The horizontal lines correspond to the basic theoretic intervals.*
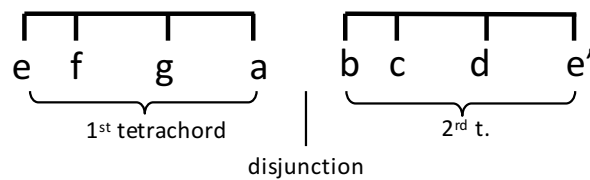

Figure 3.8: *Schema of two disjunct tetrachords, compatible with the pitch of pipes 1-8. The letter used to represent the notes does not correspond to the modern pitches.*

3.2 Designing a multimedia installation

In order to valorize the ancient Pan flute described in the previous section, a multimedia installation was developed. The valorization of an artifact located in a museum is a complex and multi-faceted task. A unique methodology does not exist, and often the design methodology is interchangeable and adaptable to the final purpose of a museum or exhibition. Some examples are the VX Principles proposed in [94] and centered on the visitor experience, or the approach based on interactive storytelling of [66].

The methodology, proposed and formalized by Silvia Gasparotto with the collaboration of the author, derives from *Design Thinking* (DT) [25], with some modifications applied to the classical approach. DT derives, in turn, from participatory design [124], co-design [125], and human centered design [50]. This approach is multi-disciplinary since it involves several professionals and their different points of view to achieve the valorization of a historical artifact. The methodology is described in the next section. Furthermore, the creation of a multimedia installation to communicate the Pan flute using the aforementioned methodology is provided, together with a summary of the assessment. A questionnaire was formulated in collaboration with Carlo Fantozzi and Edoardo Micheloni, who then gathered the answers of 23 professionals in different fields.

3.2.1 Methodology
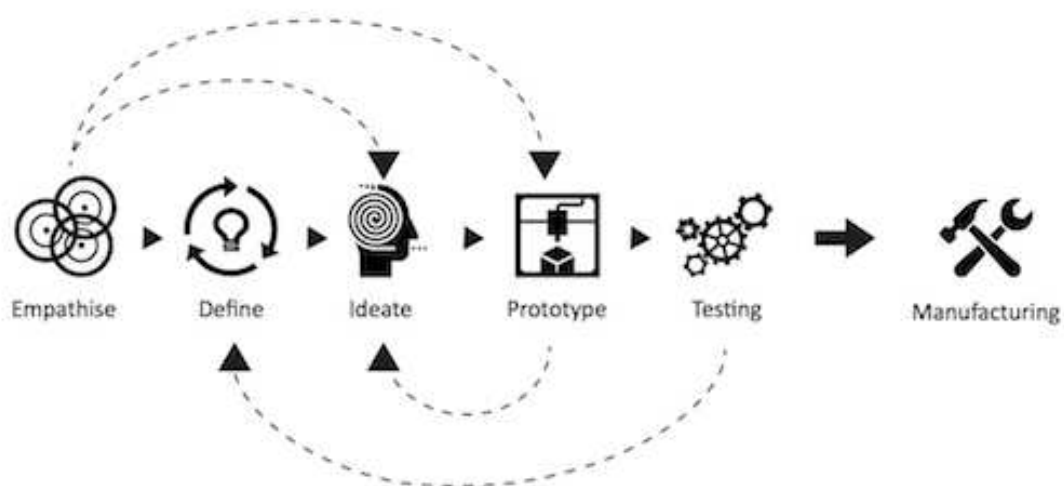
Figure 3.9: *Visual representation of the Design Thinking process.*

The DT process is mainly composed of six steps: Emphasize, Define, Ideate, Prototype, Testing and Manufacturing. As shown in Figure 3.9, the process is not necessarily linear, and it can be repeated several times before reaching the final outcome.

In the Emphasize step, the issue has to be analyzed from many points of view, in order to obtain "an empathic understanding of the people you are designing for and the problem you

are trying to solve" [52]. As for artifacts or collections in a museum or exhibition, they are, respectively, the visitors and the valorization of the artifact (or collection) itself. This phase can involve different approaches, such as brainstorming, interviews, etc. [25].
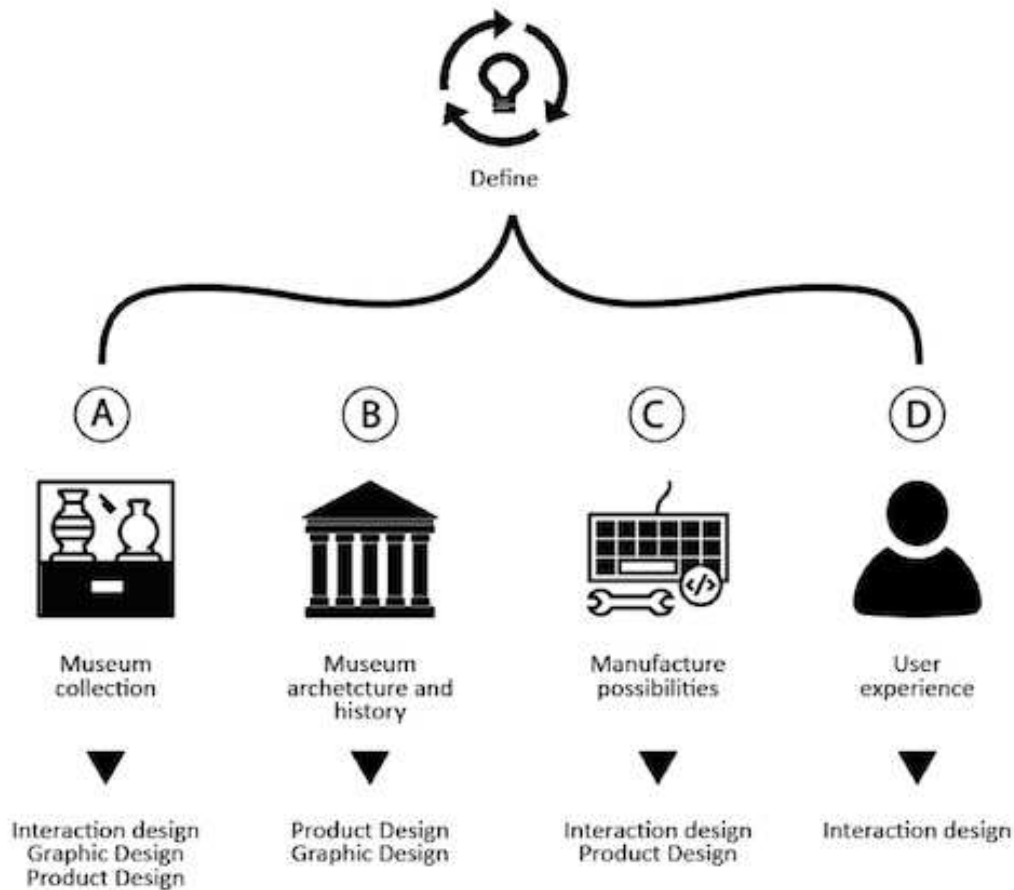
Figure 3.10: *Phases of the second step (Define).*

The aim of the second step (Define) is to delimit the "problem" or "challenge". The aspects to be considered in this analysis can be grouped in four main clusters:

A) the museum collections (or a single artifact);

B) the environment;

C) manufacture possibilities;

D) the user experience.

Figure 3.10 illustrates the four groups in a single chart.

The first phase (A) consists of an analysis of the museum collection or single artifact to valorize. The term "museum collection" has been used to be as general as possible, but it can consist of a single object. All members of the team have to be involved to identify all the characteristics of the museum collection. Furthermore, the team has to extract a subset of information or features that is important to communicate. As presented in Figure 3.10, three design disciplines are applied: Interaction Design (ID), Graphic Design (GD) and Product Design (PD). The first involves the study of appropriate kinds of interaction, such as visual, auditory, tactile, haptic, etc. The purpose of GD is to define the communication tools as well as their stylistic characteristics. Finally, PD concerns the definition of the shape of the collection's items.

In the phase (B), the museum/exhibition environment is analyzed in-depth. This means to study the space where the collection is collocated as well as the space allocated for the multimedia installation. Several aspects have to be considered such as the architectural characteristics, security issues, visiting path and more. The result is a list of rules, limitations, styles and peculiarities that characterize the environment. The two design disciplines (PD and GD) involved in this step are necessary to make sure that the aesthetics of the installation are coherent with the environment, language and global communication strategies of the museum.

In the phase (C), the aim is to realize the manufacturing and economical resources that can be allocated to the installation. The two disciplines (ID e PD) involved aims to define the development process (including possibilities of outsourcing) of the physical parts and interaction technologies suitable for the installation as well as its economical/technological limitations.

The last phase (D) of the Define section consists in analyzing the user experience. In this last phase, the visitors' points of view are considered, by envisaging how they can interact with the multimedia installation. According to [63],"in designing artifacts we do not merely design the artifacts themselves: deliberately or not, we also design conditions for their human use". Visitors differs by age, skills, knowledge, etc., thus target groups have to be considered for a reliable communication. Museum visitors are usually named as *audience*. This term refers to people who likes to watch and expects to be entertained. The aim of this phase is to transform visitors (audience) into *players* (those persons that want to have fun), and *participants* (those that do something, whether it's fun or not) [90].

After the Define step the design process continue with the Ideate step. The entire team is involved using ideation methods such brainstorming, sketching, co-creation workshop, etc. The aims is to define some ideas of installation, making also clear some features such as measures, colors, typography and interactions.

The last three stages, Prototype, Testing and Manufacturing, aim to create the final product and can be carried out in parallel. Several prototypes can be tested. The final product is the multimedia installation. Unlike industrial design, it is unique, destined to a unique museum or exhibition and is not made for series production.

3.2.2   Design of the multimedia installation

In this section, the methodology described above is applied to create a multimedia installation to communicate and valorize the ancient Pan flute presented in Section 3.1. The implementation of

each phase of the methodology is detailed in the next sections.

Empathize

In this initial phase, the entire team was involved in order to evaluate as many points of view as possible concerning the artifacts and the multimedia installation that are to be communicated. The various stakeholders presented their research and emphasized several peculiar aspects of the Pan flute. Initially, a workshop was required to understand the results of more than five years of multi-disciplinary research as well as the overall objectives of the project. Then, a brainstorming session, free of guidelines through Lateral Thinking tools [56], was conducted to collect ideas and fundamental elements that the multimedia installation should embody. A first idea of the target audience, which varies from schoolchildren to professors, also emerged from the meeting.

Define

The methodology indicated four different steps in the Define phase. Each of them is described separately. The first step consists in the study of the museum collection or the artifact. In this case, the object is the ancient Pan flute. As presented in Section 3.1, several analysis and studies were performed in several research fields:

- an archaeological study to retrace the history of the ancient musical instrument;

- a literature and iconography analysis in order to study the context and the use of the Pan flute in history;

- a set of non-invasive analysis in order to understand its components and materials as well as its current condition;

- an acoustical study with the aim to understand the tuning of the Pan flute [9].

A huge amount of information was collected, but most of them is destined to researchers. A small part of this knowledge had to be extracted in order to be communicated to the museum visitors. The only way to "understand" a musical instrument is to play it or, at least, to listen its sound. From this last statement, a first requirement can be deducted: the interaction with the multimedia installation had to include the sound of the musical instrument. Therefore, a musical interaction was required in order to communicate the instruments characteristics to the visitors and to enrich the visit. Furthermore, since the Pan flute is preserved under a glass case, a closer view can better show the peculiarities of the archaeological find.

As illustrated in Figure 3.11, part (B) concerns the architectural analysis of the museum and its collections. The Pan flute is exhibited at the Museum of Archaeological Sciences and Art placed at Palazzo Liviano in Padova. It is a XX century building designed by the designer and architect Gio Ponti. An analysis of Ponti's style was important to define the aesthetic of the multimedia installation. This has to be harmonious and integrated without compromising the surrounding environment. A survey was essential to collect measures of the area and to study existing graphic elements. The Pan flute is collocated in one of several niches, facing each other,
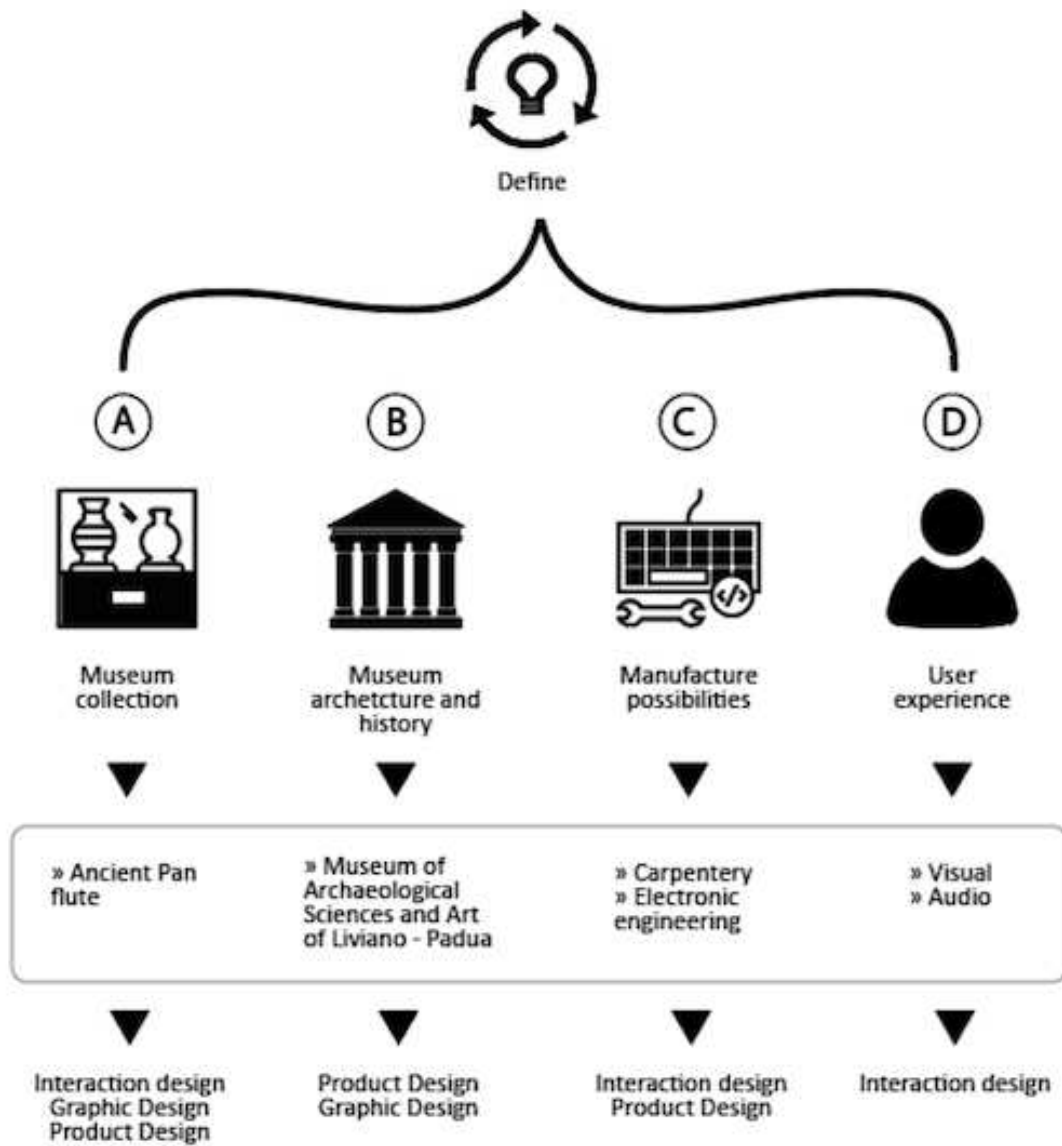
Figure 3.11: *Four stages of the Define step applied to the Pan flute project.*

of a large room with a central aisle. The multimedia installation was supposed to be collocated in one of these niches in front of the musical instrument so that visitors can easily access both the Pan flute and the installation.

From manufacture possibilities study (C), the possibility to develop the project almost completely in-house emerged. The professionals and competencies at disposal were:

- an archaeological team to manage the contents;

- a designer that developed the product, interaction and graphic design of the installation;

- a digital archaeology team to provide the 3D models and textures;

- a cartoonist for the sketches;

- a sound and music computing group able to estimate the sound, to develop the software and design hardware for the installation.

According to the analysis, only two competencies were missing:

- an expert of carpentry to fabricate the *ad hoc* forniture from the designer's project;

- an electronic company able to produce *ad hoc* electronic components.

This multidisciplinary team paved the way to a wide range of manufacturing possibilities.

The last phase (D) of this step concerns the user experience analysis. A participatory approach was required in order to present the musical instruments. The requirement emerged from this study was the necessity of an interaction with visual and auditory interaction: a visual part to describe the history, myth and research contents, and a section to reproduce the sounds of the Pan flute. In this phase, the target group (kids, ordinary museum visitors or experts) of visitors was also considered. The interaction should have different level of engagement and comprehension according to the specific target.

Ideate

The final aim was to design an interactive multimedia installation that easily explained the many aspects of the instrument, e.g. its history, myths, manufacturing and, most importantly, its sound.

During this phase, three different concepts were proposed. The first one exactly replicated the shape and the dimension of the Pan flute. This consisted in a 3D printed reconstruction with internal sensors able to detect users' blow and to reproduce the sound of the musical instrument. The second and third ideas were based on a more abstract interpretation of the musical instruments. The pan flute was stylized and represented by 14 pipes. For both ideas, an additional touch screen monitor was integrated to present the history, myth, research, etc. Figures 3.12a and 3.12b illustrate two virtual models of the two concepts.

The second concept was discarded by mutual agreement, during an internal debate, since it seemed not comfortable for visitors as well as not adequate for the architectural peculiarities of the MSA.

| (a) | (b) |

**Figure 3.12:** *The second and third concepts of the multimedia installation.*

Figure 3.13: *The main hall of the Museum of Archaeological Science and Art of the University of Padova.*

On the contrary, the third concept was more appreciated for its clean aesthetics and simple interaction. The two sections are combined in a unique configuration. On the left part, 14 cuts of different lengths represents the pipes. Above each of them, a hole, representing the mouthpiece, with an *ad hoc* sensor to capture the blow of the visitor, allows to replicate the natural interaction with the musical instrument. Historical, mythological, archaeological and technical information was instead presented through a touch screen monitor on the right part of the forniture. The aim of the design was to perpetuate the architecture, interior design and colors of the museum (Figure 3.13).

Prototype, testing and manufacturing

With rough prototyping, it is possible to immediately figure out if something works or not. Figure 3.14 shows a first prototype for the study of the interaction in the first concept. MEMs electronic components were used in order to capture the blow. Despite the size, extremely small, of these components, the space was not enough to host all of them. Furthermore, hygiene issues

definitely stopped the development. Considering the participation of a large number of visitors at the same moment (usually classes of students), it appeared impossible to guarantee sanitary safety as well as to keep the integrity of the object.
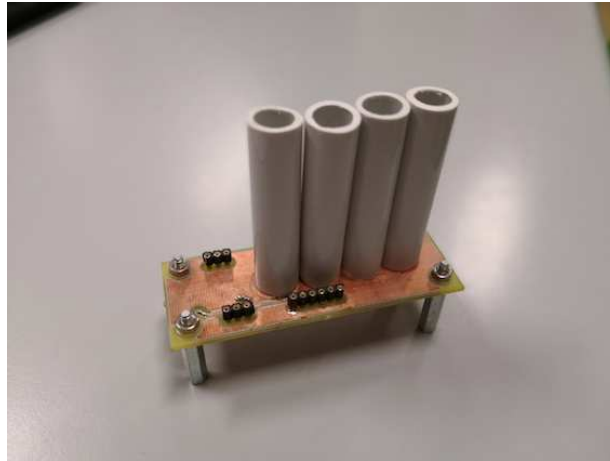
Figure 3.14: *First prototype using MEMs microphones.*

The development of the last concept was conducted by working on physical parts, interaction and graphic design in parallel. Through virtual representations, several versions of the installation were prototyped and evaluated by the team. The physical shape was realized through virtual 3D models, adjusting time after time, dimensions, and placements of the different elements.

The final multimedia installation is composed by two adjacent parts with which a visitor can interact [101]. A stylized Pan flute is carved through the oblique surface of the wooden furniture. As presented in the concept, a hole is collocated under each stylized pipe. Inside each pipe, a sensor to detect the blow of the visitors was places, whereas in each cut a LED strip was inserted in order to provide an additional visual feedback while playing the flute. The collocation of the holes (initially above the pipes) was changed during this prototyping phase, after some usability tests. In this way, the typical interaction with this kind of musical instrument was simulated. The main technological issues of the virtual instrument concern:

- the acquisition of the blow's signal;

- recognition of events that can define its behavior;

- synthesis of the sound of each pipe.

Although the technologies developed for the project involve *ad hoc* components, derived from the project specifications, the architecture of the system can be used and easily adapted to other projects whose aim to translate or interpret the behavior of a blow.

In order to create the virtual musical instrument, two main synthesis techniques were proposed: a physical model or a wavetable [49]. Although the first option is more accurate, a correct and complete implementation is more complicated and requires a different set of inputs. For example, it is not possible to extract the direction of the blow to reproduce the virtual inclination of

the Pan flute or to easily detect the minimum air pressure needed to trigger the natural resonance of a pipe. On the contrary, the second model is easier to be implemented and it provides a good reliability of the original sound's synthesis. Furthermore, the aim of the installation should be considered: enabling to listen and learn the sound of the instrument and the way it was played. In conclusion, this analysis confirmed the second model. A set of *ad hoc* sensors was connected to an Arduino Mega 2560[3] and mounted in one furniture's slot (see Figure 3.15).

The envelope function of a sound is a smooth curve outlining its extremes [84] that can be segmented in four main sections: attack, decay, sustain and release. In the signal obtained recording a blow by means of a microphone, it is possible to identify such segments. In the MIDI protocol, they can be represented in three messages: *NoteOn* (attack), *ControlChange* (decay, sustain), *NoteOff* (release). With the first message, the sampled sound of a pipe starts its reproduction. The second one provides a way to vary the interpretation of the blow, whereas the latter interrupts the playback. Sampled sounds of a generic Pan flute were tuned in order to achieve the estimated fundamental frequencies of the pipes. The samples and the MIDI messages are managed through Ableton Live 9[4]. Then, each sample was computed in order to emphasize both the attack of the note (with the usual "T" sound) and the high frequency (the effect of the air blew in the pipe), resulting in a more realistic sound.

Furthermore, a second fluttered sample has been added to each note according to the pressure of the blow, in order to accentuate the difference of intensity and the dynamic of the sound.
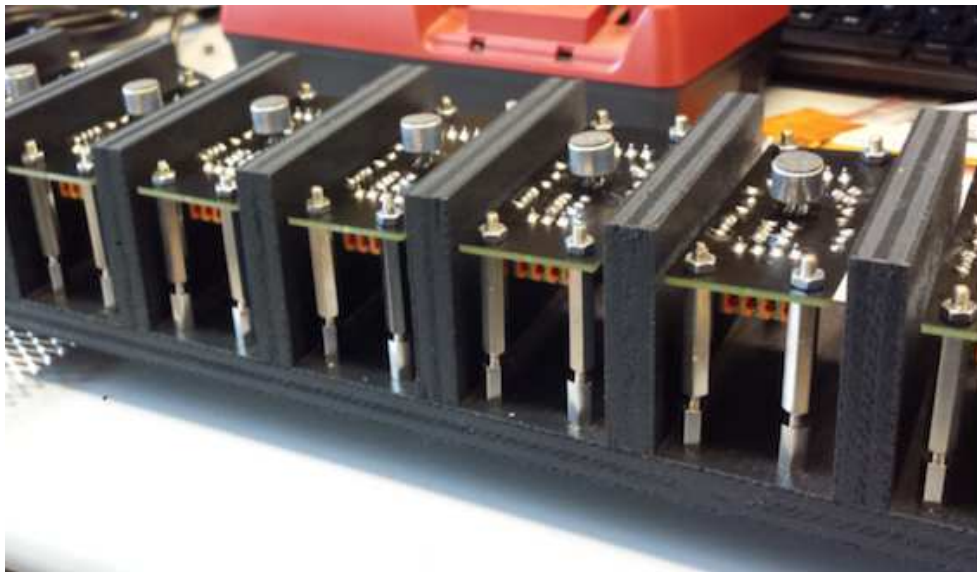
Figure 3.15: *PCB of the microphone sensors to detect visitors' blow.*

The second part of the multimedia installation consists of a 42" touch screen, connected to a pc, embedded in the furniture. The application is developed using the framework Unity 3D[5].

---

[3]`arduino.cc/en/Main/ArduinoBoardMega2560?setlang=en` (Retrieved September 24, 2018)

[4]`ableton.com/en/live/` (Retrieved September 24, 2018

[5]`unity3d.com/` (Retrieved September 24, 2018)

The information is subdivided in five different chapters reachable through a menu (Figure D.1): Myth, History, Sound, Flute, and 3D. All sections have common elements, such as the font (Titillium), the chromatic range and the menu bar at the bottom of each screen, whereas some other interchangeable features are customized on the basis of contents.

In the first section (Figure D.2), several comic strips narrate the mythological story of the Pan flute. Through simple swipes, the visitors can explore the excerpts of Ovid's Metamorphoses and their related illustration.

The section Sound (Figure D.3) provides an alternative interaction with the virtual Pan flute. A stylized musical instrument is visualized and can be played by simply touching its pipes. The physical and virtual sound parts are developed by mutual exclusion, so it is impossible to play both at the same time. Unlike the virtual instrument that uses the blow as input, the touch version does not allow to change the sound level.

The section History (Figure D.4) proposes a meaningful part of the European literary and iconographic sources collected by archaeologists involved in the project. Each source is disposed in a temporal bar that covers the range between VIII century B.C. and the X century A.D. By selecting a mark of the desired source, a pop-up window shows images or texts. Each source is also geographically located on a map of Europe.

The Pan Flute section is composed by four thematic subsections:

- the archaeological information about the discover of the Pan flute (Figure D.5a);

- the cultural and musical context in which the instrument was produced (Figure D.5b);

- the study concerning the sound of the musical instrument (Figure D.6b);

- the reconstruction of a similar copy of the archaeological find applying experimental archaeology[6] (Figure D.6a).

The last section consists of two parts concerning the virtual 3D model and the Computerized Tomography (CT).

The raw model, presented in Figure 3.1.1, was fulfilled using a texture obtained from several photos. With well-known gestures on the touchscreen, such as swipe and pitch-to-zoom, it is possible to rotate and thoroughly examine the exterior peculiarities of the musical instrument. The results are illustrated in Figure D.7a.

The second part (Figure D.7b) provides a different approach to the exploration of the CT. A set of markers placed on a bar enables to discover all the peculiarities of the artifact. With a simple click on the marker, the CT browses the three orthogonal views. As soon as the selected peculiarity is shown and highlighted, a description appears on the left of the CT model.

The resulting multimedia installation can be observed in Figure 3.16. It is located in the museum, where it is now permanently exhibited. After a while, a new re-collocation of the multimedia installation was required. With a large number of visitors, the proximity of the artifact constitutes a danger for the artifact itself. The next section reports the assessment of the methodology and shows the impact of this choice on the overall evaluation.

---

[6]realized by the experimental archaeologist Mauro Cesaretto

Figure 3.16: *The multimedia installation, permanently exhibited at the Museum of Archaeological Science and Art of the University of Padova.*

3.2.3 Assessment

As mentioned in the introduction of Section 3.2, a questionnaire [10, 95] was prepared in order to assess the user experience as well as the decisions taken during the design phase (Section 3.2.2), which follows the methodology described in Section 3.2.1. The assessment was proposed to 23 professionals over a period of 7 weeks. The consulted group of professionals included researchers and professionals with experience in one or more of these fields: engineering, music, musicology, history. The decision to not involve the generic public stems from the fact that the assessment focus was on the methodological and technical aspects of the multimedia installation. The experts' average age was of about 52 years. Only three of them were under 40 years old whereas six were more than 60 years old. Most of them were university professors (16). The remaining ones were researchers or freelance professionals.

The questionnaire was divided in two distinct parts: one concerning the user experience, and one about the methodology and design.

The assessment procedure consists of the following sequence of steps. In the first step, the subject was allowed to observe the real artifacts and interact with the multimedia installation. The initial information at his disposal was minimal: they knew that the interaction through the panel was possible by touch or blowing only. Then, the expert was left alone, as he/she was visiting the museum by herself/himself. The interaction time was recorded, without the subjects knowing. The average duration was 16 minutes. At the end of the interaction, the first part of the questionnaire (19 questions) concerning the user experience was compiled by each expert. The

split of the questionnaire was required as the user experience involves a significant amount of emotions that could be spoiled by rational communication.

After this first part, an open discussion session was held. The design methodology was illustrated and the objectives of the research were openly stated. Moreover, pending questions from experts concerning the original instrument, the multimedia installation, technologies, etc. were answered. Sometimes, this explanation was directly given by using the multimedia installation.

After the discussion, each subject answered to the second part of the questionnaire (25 questions). This concerns the three main phases described in the define phase of the methodology: the museum collection (in this case, a single artifact), the environment (the museum) and the manufacture possibilities.

All the answers of the questionnaires were based on the Likert method: for each statement, the subject has to indicate how much she/he agreed on a 5-level scale from 1 (strongly disagree) to 5 (strongly agree). A detailed description of the questionnaires is postponed to a further publication. A summary of the assessment is here reported.

From the quantitative results of the assessment, we can say that, all in all, the experts liked the interaction with the installation, they appreciated our methodology, and they judged our aims to be fulfilled by the installation.

Concerning the first part of the questionnaire, 15 out of 19 statements received an average score above 4, meaning that the experts largely appreciated the user experience. The higher votes were bestowed to the simplicity and easiness to understand the touch screen part, as well as for easiness of navigation structure. An important result concerns the interactive experience, which most of the experts considered as pleasant. Only four statements concerning the blowing interaction received a score below 4. Here the statements are reported:

4. BlowFlute is a convenient means of interaction (average vote: 3.61).

5. BlowFlute is simpler to use than the touch screen to appreciate the flute sound (average vote: 2.78).

18. The variation over time of the flute sound was perceptible (average vote: 3.26).

19. I easily perceived the nuances related to the sound of the flute (average vote: 3.43).

The first two are easily explainable, because it is effectively easier to tap a pipe then blowing into holes. Some of the experts also criticized the height of the holes, and the necessity to stoop. It must be remarked that some compromises were necessary in order to satisfy all the requirements. For example, it was essential to maintain the installation sober and visually integrated into the museum, without no prominent cues, to safeguard the visitors from hygienic threats as well as considering the usability for both kids and adults. Concerning the last statements, the variation is not explained and a high level of attention is required in order to perceive the effects.

The second part also returned good results, with 21 out of 25 statements scoring above 4. It can be stated that the installation effectively presents the information about the flute, better then conventional means. The possibility to manipulate the artifact received a great appreciation. The possibility to interact with the sound was especially regarded as a plus.

The only four statements with lower scores are:

20. The installation integrates aesthetically in the context of the room where it is located (avg: 3.57).

21. The installation aesthetically enriches the room where it is located (avg: 3.00).

35. Manipulating a virtual model of the flute is better than manipulating a physical reconstruction of the flute (avg: 3.39).

37. Blowing into a hole (BlowFlute) is preferable with respect to other blowing possibilities (e.g., blowing into a straw) (avg: 3.87).

The scores of the first two statements were heavily influenced by the fact that the installation is not currently placed in the exact museum spot that was originally designed for. Therefore, the design connections with the original room are broken. The third statement is a critic pointed towards the constraints of the museum. A physical reconstruction of the flute is too fragile for kids as well as too difficult to clean when handled by hundreds of visitors.

In conclusion, the experts assessment shows that such design decisions, albeit sensible, had a measurable impact on the appeal of some aspects of the installation.

Chapter 4

Conclusions

4.1 Conclusion

Musical heritage is a huge and multi-faceted part of the cultural heritage. By applying information engineering, sound and music computing research field can help to delineate methodologies to create, preserve, access and experience the digital version of this prominent heritage. This thesis fits on this research field and proposes new methodologies for exploiting, valorizing and experiencing musical heritage into the digital domain, by maintaining the relation with its cultural context. The digitization opens several opportunities, such as to present easier ways to analyze, promote, spread and valorize a cultural object. At the same time, it implicates issues concerning the authenticity of the digital artifact. Three paradigmatic case studies are selected: tape music, an Arab-Andalusian corpus and an ancient Pan flute. Such case studies were selected as they well represent the opportunities which information engineering can enable in the cultural heritage fields. The methodologies, algorithms, experiment, results and outcomes, as well as further work prepared for each case study are presented in the next sections.

4.1.1 Case study: tape music

As presented in Section 2.1, tape music is a paradigmatic musical genre which revolutionized the music creation process, becoming one of the most important cultural phenomena since the 1950s. The creative process of this kind of music involved the physical manipulation of the tape, which in turn exacerbates the issues related to its preservation, accessing and experiencing. These philological issues can also be extended to other music genres: since the 1960s, these working methods based on cuts, editing, as well as on synthesized and "concrete" sounds, has been used in rock, jazz, for movie sound tracks, as well as for musical work as *Prometeo* by Luigi Nono and the musical part of multimedia theatrical works such as *Medea* by Adriano Guarnieri.

After defining the issues concerning tape music works, its contextual information and the concept of discontinuity, a set of methodologies and a software for the automatic analysis of the digitized audio documents is presented. This was developed by using machine learning, computer vision and neural networks techniques.

First of all, several classes of discontinuities were detailed. Then, an approach to the discontinuity detection in the video of the tape is presented. A dataset of discontinuities was realized by using neural networks and will be the base of further experiments. These will probably involve the same technology to classify the detected discontinuities. This kind of tools aims to relief audio technicians and musicologists involved, respectively, in the digitization process and the musicological analysis, from repetitive tasks. As for the scholar point of view, automatic tools that advise discontinuities into the audio documents represent a potentially powerful aid for a correct and complete musicological analysis.

Concerning equalization, the proposed experiment proves the suitability of machine learning techniques for the automatic analysis of equalization through the analysis of the background noise of a digitized audio recording. This kind of tool can assist audio technicians in technical decisions related to the replay device configuration, improving the faithfulness of the digital preservation copy. Furthermore, this tool provides a reliable method to verify the correctness of the digitization of other works and, if necessary, compensates (only in an access copy) the error, obtaining the original sound. The excellent results lay in the foundation for future developments that go beyond electro-acoustic documents. In general, to take advantage from the outputs of algorithms, software for automatic quality control and validation kits must be developed in order to aid digitization operators and analysis tools for scholars able to provide and manage discontinuities and points of interest synchronized with the audio and video.

Another important work presented in this thesis is related to stemmatics and the application of phylogenetic and computer vision algorithms to automatize the philological analysis. Phylogenetic analysis of tape music aims to reconstruct the generation history of different copies of the same audio content. This kind of innovative approach was applied for the first time in this field. The proposed solution presents a satisfying robustness to the adoption of the wrong reading setup (i.e., with speed, equalization and filtering different from those adopted in the creation of the tape), together with a good reconstruction accuracy of the phylogenetic tree. The reconstructed dependencies proved to be correct or plausible in 90% of the experimental cases. This work paves the way to a further extension of the proposed approach, in which a widened set of editing techniques, intentional and unintentional alterations, configurations, as well as different tape recorders and syndromes, can be included.

With respect to the experiencing of digitized audio documents, this work proposes a novel philologically informed methodology for accessing this kind of recordings which considers their cultural context. Based on such methodology, an Android app which virtualizes a tape recorder is presented with its assessment. Furthermore, three other web based applications are detailed. Two of them virtualize a gramophone and a tape recorder, respectively. They propose a set of tools in order to configure the main parameters (equalization, speed, etc.) of these analog replay machines. The last web interface is a prototype of an analysis tool for those musicologists that exploit contextual information and the results of the discontinues detection tool. These contributions advance the state of the art concerning the access to historical audio documents, by taking advantage of web and mobile technologies. Furthermore, they lay the foundation for an ambitious project concerning an integrated environment for preserving, analyzing, accessing and experiencing digitized audio documents, namely the Access Kit. The main parts of this project, presented in Figure 4.1, consists of a distributed system able to store and manage the preservation

and access copies of most types of analog audio recordings. This system is based on a central server unit. This enables to manage all the requests of the web, desktop and mobile interfaces utilized for the selective access of digitized audio recordings, and to execute the automatic analysis detailed in this work.
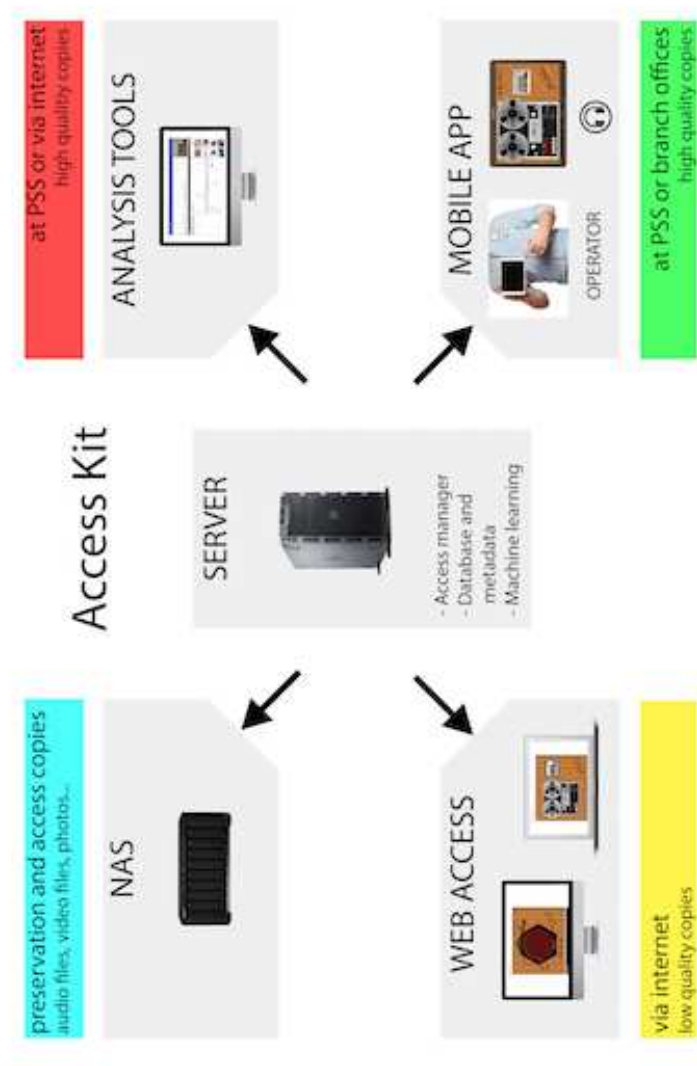
Figure 4.1: *Schema of Access Kit, an integration of all the analysis and accessing tools presented in the thesis.*

4.1.2   Case study: Arab-Andalusian corpus

Arab-Andalusian music is a unique music tradition, which reflects the rich culture developed in the encounter of Western and Eastern Mediterranean cultures in the Iberian Peninsula [37]. The Arab-Andalusian music research corpus, presented in Section 2.7, offers a large collection of openly accessible, well curated and annotated data (audio recordings, scores and lyrics), which is presented with the aim of fostering the research of this music tradition, both from the musicology and MIR fields.

Furthermore, freely accessible Jupyter notebooks were developed for computational musicologists who want to browse and retrieve the corpus data and metadata, and to exploit its potentiality through a data-driven approach. The first algorithm of computational musicology applied to Arab-Andalusian music is presented, alongside the dataset and the experiments applied to validate its suitability. The algorithm supports the *nawba* recognition task by matching templates computed from music scores with the pitch distributions extracted from the audio recordings. The experiment to evaluate the algorithm provides a good performance, with an accuracy of 75%. This is only a first step for the analysis of this important music tradition.

From the musicological point of view, the three collections of data and related metadata and annotations offer a great potential for the analysis of the melodic and rhythmic elements of this music tradition. These machine-readable data and metadata enable the computation of quantitative and statistical information, regarding pitch, intervals, predominant scales degrees, intonation profiles, note durations, meters, tempograms, etc.

From a MIR point of view, it also suggests interesting research tasks, such as the predominant melodic extraction from a heterophonic source, the fundamental degree detection, *nawba*, *ṭābʿ* or *mīzān* automatic identification and classification, structural segmentation, tempo and meter estimation, etc. The manual annotations provide a valuable ground-truth for these tasks. Further works can deepen audio to score or audio to lyric alignments, adapting existing algorithm created during the CompMusic project [121, 61]. The strict relationship between the lyrics content and melodic features is a promising field of study in Natural Language Processing methods. This relation can be studied, as in [152], in order to understand how melodic and rhythmic modes are selected according to the lyrics contents.

4.1.3   Case study: Archaeology and virtual acoustics

The multidisciplinary work concerning the valorization of the ancient Pan flute described in Section 3.1 brings important outcomes. The first one is a methodology to develop a multimedia installation that communicates and valorizes archaeological musical instruments by considering the cultural context. The multimedia installations can be a valid mean to provide an interaction with an artifact, that is usually not touchable or playable in museums. The interaction model utilized to provide access to the general public leverages on a multi-sensory (visual, auditory, tactile) interplay that includes both contextual information and a virtual counterpart of the artifact. The methodology includes an adaptation of Design Thinking to the context of interactive museum installations, and a matching design process that is deeply interdisciplinary.

The proposed methodology was applied to develop a multimedia installation for an ancient

Pan flute exhibited at the Museum of Archaeological Science and Art of the University of Padova. A complete description of the design process and a final installation is proposed by highlighting the importance of each phase and, at the same time, elucidate its statements. This methodology as well as the practical findings can be fruitfully applied to other museum artifacts and in a varied range of cultural contexts.

A group of experts evaluated this installation and the methodology through a questionnaire. The assessment results, presented in Section 3.2.3 have confirmed that the methodology has led to a user-engaging communication of different aspects related to the history, iconography, and acoustics of the Pan flute. Further improvements are however possible in the interaction through the sound, considering in any case the constraints of the museum. Moreover, other studies on tuning can be performed, merging the musicological theory in order to provide a better approximation of the original sound. Otherwise, a new virtual instrument based on its physical model can improve the presented one.

Appendix A

Frames and photos of Discontinuities

This appendix contains several discontinuity frames that are part of the datasets. Examples of each class are presented in the following figures.

Figure A.1: *A frame of the class End of Tape: the capstan is detached from the tape.*

Figure A.2: *A frame of the class Shadows.*

Figure A.3: *A frame of the class Ripples.*

(a)

(b)

(c)

(d)

Figure A.4: *Four frames of the class Leader Tape / Magnetic Tape Splices.*

(a)

(b)

(c)

(d)

Figure A.5: *Four frames of the class Magnetic Tape / Magnetic Tape Splices.*

(a)

(b)

Figure A.6: *Two photos of brands printed on the back of the tape.*

(a)

(b)

(c)

(d)

Figure A.7: *Four frames of the class Brands.*

(a)

(b)

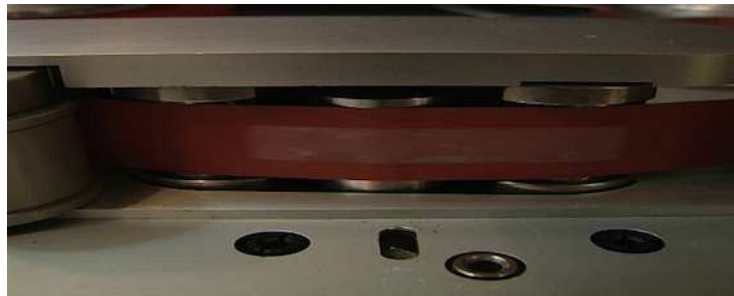**Figure A.8:** *Two frames of the class Damaged Tape.*

(a)

(b)

Figure A.9: *Two frames of the class Dirt.*

(a)

(b)

Figure A.10: *Two frames of the class Marks.*

Appendix B

Assessment of REMIND app.

This appendix contains supplementary material about the assessment of the REMIND app [67] (see Section 2.6.2). The assessment was based on a set of 14 questions, which are reported below.

Q1. Does the app always present information in a clear fashion?

Q2. Are text messages and icons clear and unambiguous?

Q3. Is the navigation structure (that is, how information is divided into several screens) easy to remember?

Q4. Is the app sufficiently self-explanatory?

Q5. Is the touch screen a convenient mean of interaction?

Q6. Is the skeuomorphic interface actually interesting, or do you consider it a secondary detail?

Q7. Do you think it is correct to offer no way of accessing music beyond sequential access (e.g., no random access)?

Q8. Are the dimensions of the tablet adequate?

Q9. Were all tasks easy (that is, no task that was thought to be easy proved difficult in practice)?

Q10. Are metadata complete (that is, no useful metadata is missing)?

Q11. Is the app complete (that is, no useful function is missing)?

Q12. Are all implemented features actually useful?

Q13. All in all, do you believe the app would allow you to analyze a piece of electronic music faster?

Q14. All in all, do you believe the app would allow you to analyze a piece of electronic music more accurately?

Full answers from the professionals are available in Table B.1.

Table B.1: *The full quantitative answers for the assessment of the REMIND app [67]. Questions were answered on a 5-level scale from 0 (most negative opinion) to 4 (most positive opinion). Professionals are identified with the labels "P1"-"P20" to keep them anonymous.*

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
P1	3	3	3	3	3	1	0	3	4	4	0	4	1	1
P2	3	3	4	3	4	3	0	4	4	1	1	4	2	3
P3	4	4	4	4	4	1	0	4	4	4	0	4	1	1
P4	4	4	4	4	4	4	0	4	4	4	3	4	3	4
P5	4	4	4	4	4	4	0	4	4	2	2	4	3	4
P6	4	3	4	3	4	3	0	4	4	1	3	4	2	3
P7	4	4	4	4	4	4	0	4	4	4	4	4	0	4
P8	3	3	3	3	3	3	1	2	4	2	1	4	0	3
P9	4	4	4	4	4	3	0	4	4	1	2	4	3	4
P10	4	3	4	3	4	3	0	4	4	1	1	4	2	4
P11	3	3	4	3	4	3	0	4	4	1	1	4	2	3
P12	4	3	4	3	4	3	1	4	4	1	1	4	2	4
P13	3	3	3	4	3	3	0	4	4	1	0	4	2	3
P14	4	4	4	4	4	3	0	4	4	1	2	4	2	3
P15	4	3	4	3	3	4	0	4	4	1	2	4	1	3
P16	4	3	4	3	3	4	0	4	4	1	2	4	2	4
P17	4	4	4	4	4	4	1	3	4	2	3	4	3	4
P18	3	3	4	3	4	4	1	4	4	1	1	4	2	2
P19	3	3	3	3	3	3	1	4	4	1	1	4	1	3
P20	4	4	4	4	4	4	1	4	4	1	1	4	2	3

Appendix C

Arab Andalusian corpus statistics

This appendix contains all the statistics concerning the four musical entities described in Section 2.7: *nawabāt*, *ṭubū'*, forms and *mawāzīn*. For each entity, a table shows original Arabic name, the transliterated name, the number of sections or recordings and the overall section time of each element. Only the typologies covered by the corpus are included in the tables.

Table C.1: *The list of nawabāt of the Arab-Andalusian corpus. The noncanonical one is marked with *.*

Name	Transliterated name	Recordings	Duration
الاستهلال	al-istihlāl	22	17h 00' 45"
الاصبهان	al-iṣbahān	13	9h 27' 14"
الحجاز الكبير	al-ḥiŷāz al-kabīr	7	5h 21' 57"
الحجاز المشرقي	al-ḥiŷāz al-mašriqī	14	7h 12' 47"
الرصد	al-raṣd	10	7h 01' 57"
العشاق	al-ʿuššāq	7	4h 35' 26"
المالة	al-māya	12	9h 16' 06"
رصد الذيل	raṣd al-ḏāyl	17	12h 08' 18"
رمل المالة	raml al-māya	16	12h 43' 54"
عراق العجم	ʿirāq al-ʿaŷam	7	4h 08' 54"
غريبة الحسين	garībat al-ḥusayn	11	6h 54' 20"
بواكر المالة	bawākir al-māya *	2	1h 23' 16"
بدون	none	21	2h 56' 07"

Table C.2: *The list of ṭubūʿ of the Arab-Andalusian corpus. Noncanonical ones are marked with *.*

Name	Transliterated name	Secs.	Duration
الاستهلال	al-istihlāl	146	17h 10' 34"
الاصبهان	al-iṣbahān	72	9h 38' 37"
الحجاز الكبير	al-ḥiŷāz al-kabīr	49	5h 41' 10"
الحجاز المشرقي	al-ḥiŷāz al-mašriqī	25	3h 09' 51"
الرصد	al-raṣd	55	7h 01' 57"
العشاق	al-ʿuššāq	40	4h 35' 26"
المالة	al-māya	73	9h 16' 06"
رصد الذيل	raṣd al-ḏāyl	89	12h 08' 18"
رمل المالة	raml al-māya	95	13h 11' 37"
عراق العجم	ʿirāq al-ʿaŷam	41	4h 08' 54"
غريبة الحسين	garībat al-ḥusayn	77	7h 07' 09"
بواكر المالة	bawākir al-māya *	11	1h 23' 16"
الصيكة	al-ṣīka	1	7' 11"
المشرقي	al-mašriquī	57	4h 08' 40"
الزركة	al-zerga *	6	31' 59"
قريب إلى الاستهلال	similar to al-istihlāl *	2	9' 41"
قريب إلى المالة	similar to al-māya *	2	11' 05"
رصد الذيل مغربي	Moroccan raṣd al-ḏāyl *	2	8' 29"
خليط الطبوع	mixed ṭubūʿ	1	7' 03"
النهاوند	Nahawand (maqam) *	3	13' 58"

Table C.3: *The list of forms of the Arab-Andalusian corpus.*

Name	Transliterated name	Sections	Duration
إنشاد	*inšād*	75	2h 43' 52"
توشية	*tawāšī*	105	3h 06' 10"
مشالية	*mišālia*	168	4h 44' 06"
تقسيم	*taqsīm*	22	43' 54"
موال	*mawwāl*	16	1h 27' 12"
موسع	*muassaʿ*	138	46h 08' 07"
مهزوز	*mahzūz*	144	14h 37' 25"
انصراف	*inṣirāf*	178	25h 34' 58"
صامت رصد الذيل	Instr. *raṣd al-ḏāyl*	1	5' 17"

Table C.4: *The list of mawāzīn of the Arab-Andalusian corpus. Non-canonical meters are marked with *.*

Name	Transliterated name	Sections	Duration
بسيط	*basīṭ*	105	22h 16' 08"
بطايحي	*bṭāyḥī*	115	23h 30' 45"
درج	*darŷ*	17	2h 42' 26"
قائم ونصف	*qāʾim wa niṣf*	97	14h 56' 07"
قدام	*quddām*	174	23h 46' 39"
ثنائي	binary *	50	2h 10' 25"
ثلاثي	ternary *	2	6' 09"
حضّاري	*ḥaddārī* *	1	10' 24"
ثنائي وبطايحي	binary and *bṭāyḥī* *	1	5' 17"
زنداني وقدام	*zandānī* and *quddām* *	5	50' 42"
بدون	none	280	9h 35' 59"

Appendix D

Screenshots of the multimedia installation

The multimedia installation described in Section 3.2 has a touch screen based application to communicate several information about the ancient Pan flute presented in Section 3.1. In this Appendix, Several screenshots illustrate the sections of multimedia installations.

Figure D.1: *The main menu of the multimedia installation.*
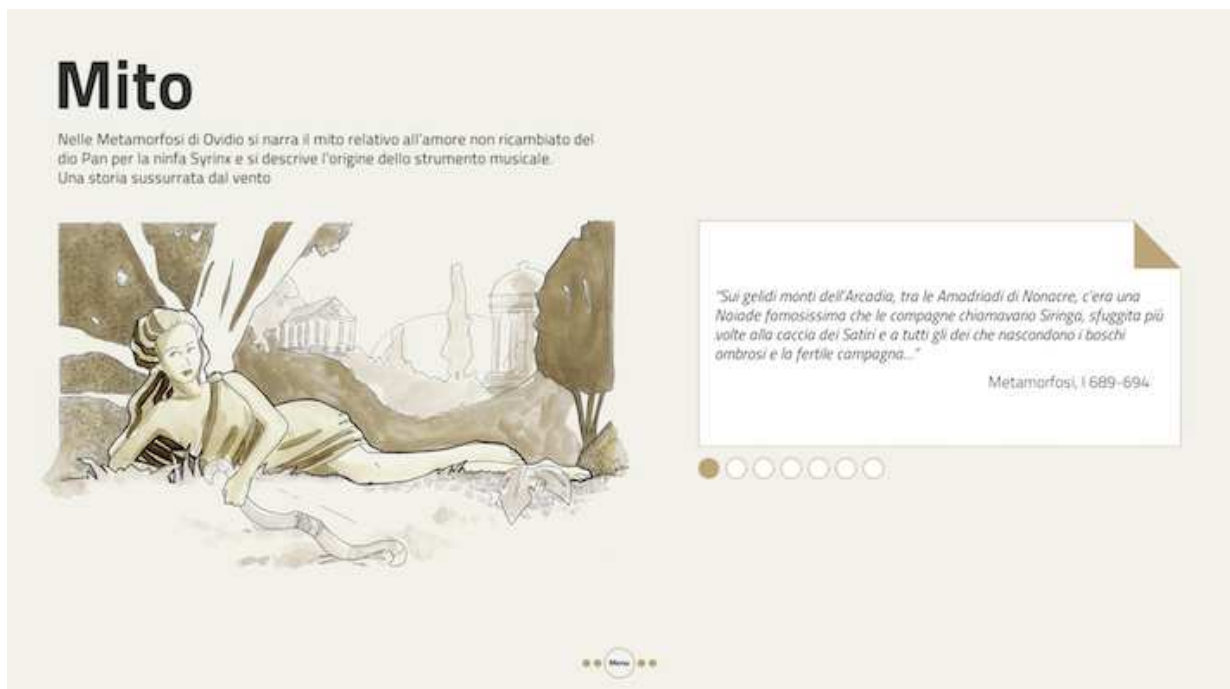
Figure D.2: *The Myth section of the multimedia installation, with comic strips and Ovid's Metamorphosis excerpts.*
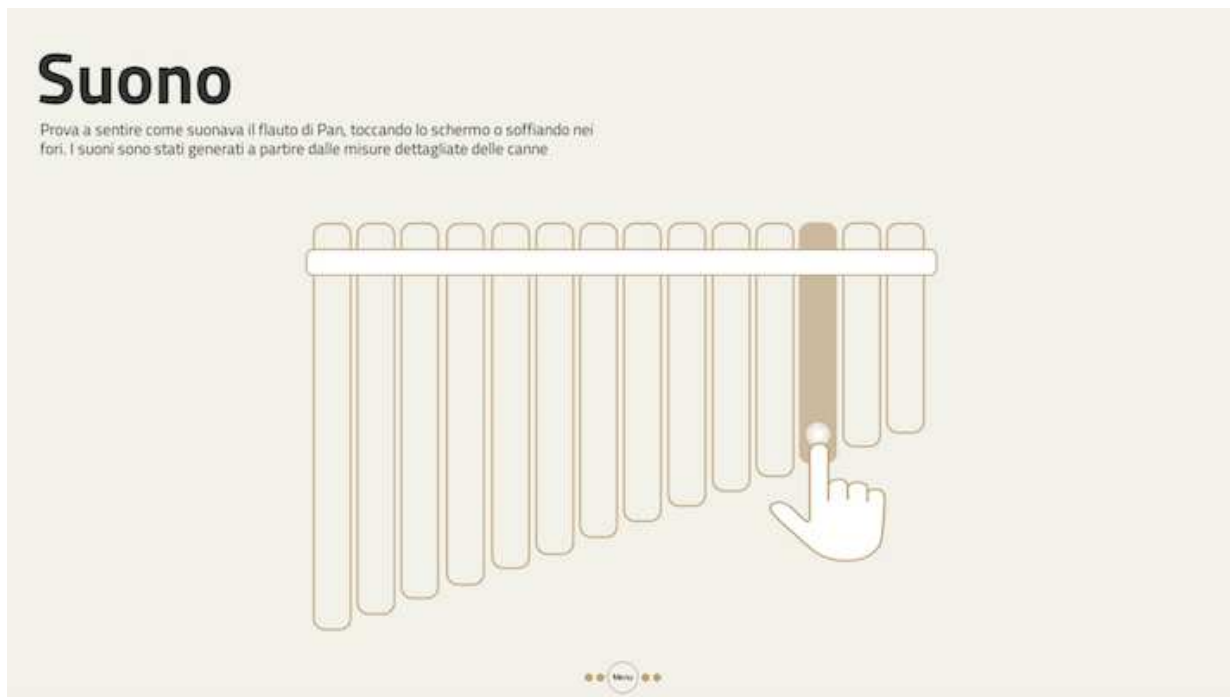
Figure D.3: *The Sound section of the multimedia installation, where it is possible to play the virtual instrument simply touching a pipe.*
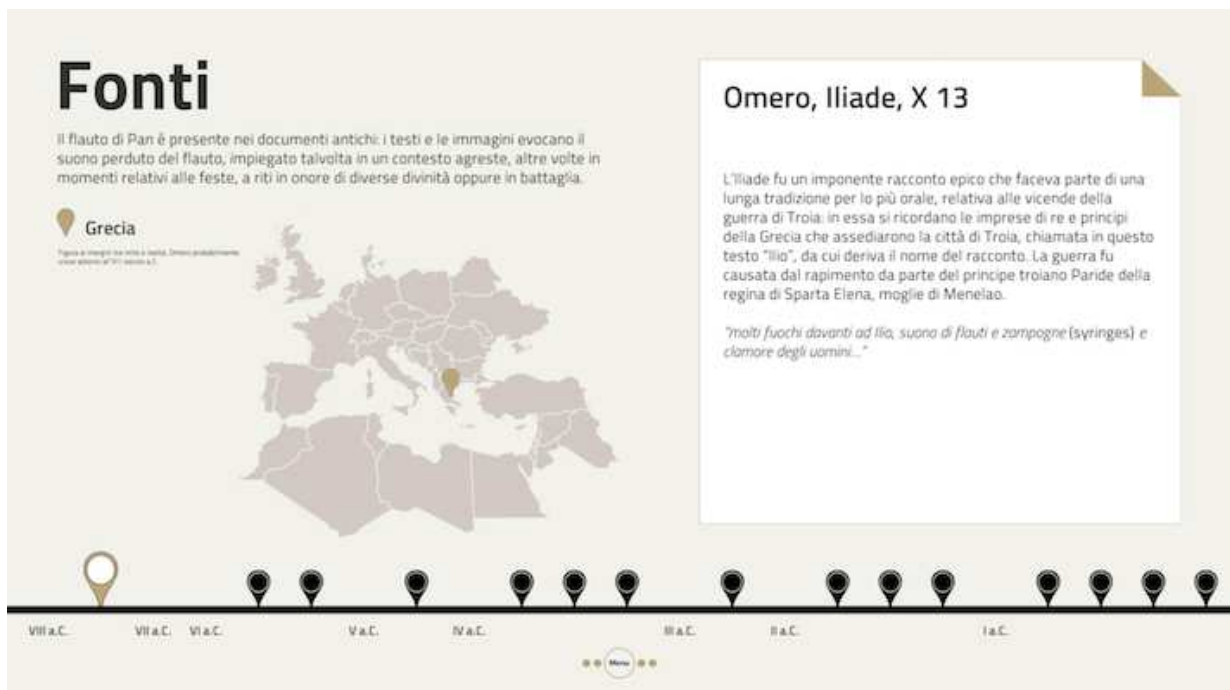
Figure D.4: *The History section of the multimedia installation, with meaningful part of the European literary and iconographic sources.*

(a)

(b)

Figure D.5: *The first two parts of the Pan Flute section of the multimedia installation, with information concerning the cultural and musical context (Figure D.5b) and history of the ancient Pan flute (Figure D.5a).*
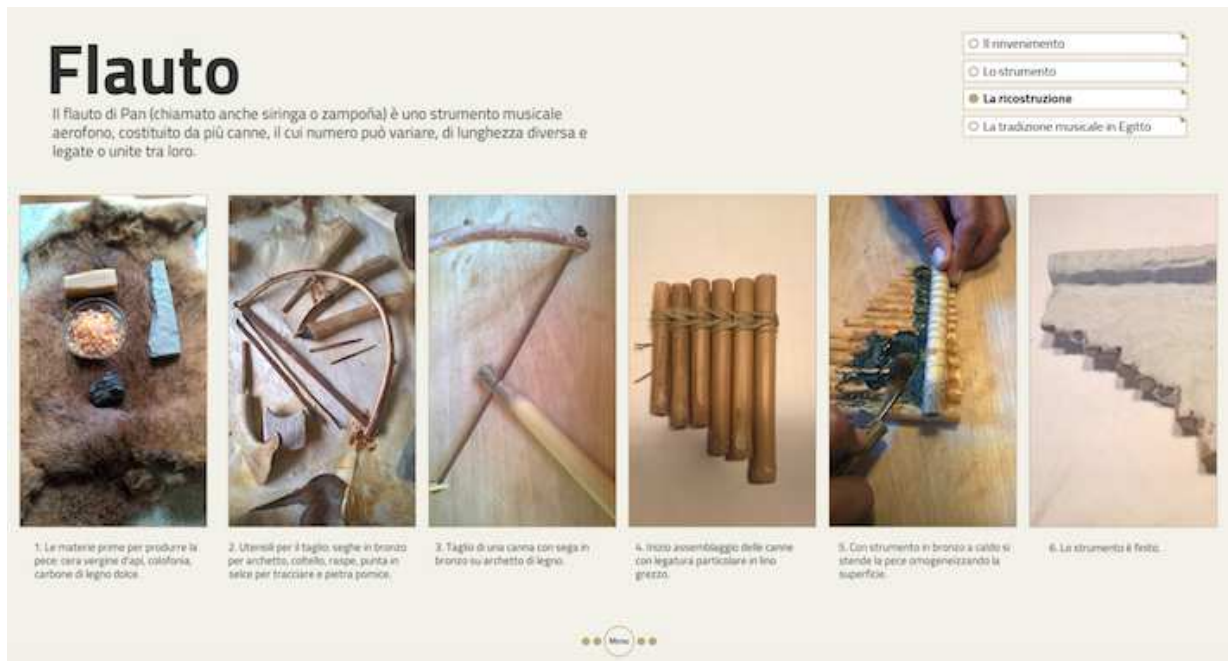
(a)

(b)

Figure D.6: *The last two parts of Pan Flute section, with information concerning the reconstruction of a copy applying experimental archaeology (Figure D.6a) and tuning research (Figure D.6b).*

(a)

(b)

Figure D.7: *The 3D section of the multimedia installation has two parts: one with the 3D scan (Figure D.7a) and one with the Computerized Tomography scan (Figure D.7b).*

Bibliography

[1] P. Adenot and R. Toy. Web audio API, W3C working draft, 19 june 2018. Technical report, W3C, December 2018.

[2] T. W. Adorno and R. Hullot-Kentor. *Philosophy of new music*. MLA, 2006.

[3] AES. *AES recommended practice for audio preservation and restoration – Storage and handling – Storage of polyester-base magnetic tape*. AES, New York, NY, 1997 (r2008).

[4] Arturia. iMini. http://www.arturia.com/products/ipad-synths/imini/overview, 2014.

[5] H. S. Atlı, B. Uyar, S. Sentürk, B. Bozkurt, and X. Serra. Audio feature extraction for exploring Turkish makam music. In *3rd International Conference on Audio Technologies for Music and Media*, pages 142–153, 2014.

[6] H. S. Atlı, B. Bozkurt, and S. Şentürk. A method for tonic frequency identification of turkish makam music recordings. In *5th International Workshop on Folk Music Analysis (FMA 2015)*, pages 119–122, Paris, France, 2015.

[7] H. S. Atlı, S. Sentürk, B. Bozkurt, and X. Serra. Synthesis of Turkish makam music scores using an adaptive tuning approach. In *7th International Workshop on Folk Music Analysis (FMA)*, pages 66–71, 2017.

[8] F. Avanzini, S. Canazza, G. De Poli, C. Fantozzi, N. Pretto, A. Rodà, I. Angelini, C. Bettineschi, G. Deotto, E. Faresin, A. Menegazzi, G. Molin, G. Salemi, and P. Zanovello. Archaelogy and virtual acoustics. a pan flute from ancient egypt. In *Proceedings of 12th International Sound and Music Computing Conference (SMC'15)*, pages 31–36, Maynooth, Ireland, August 2015. doi: 10.5281/zenodo.851067.

[9] F. Avanzini, S. Canazza, G. De Poli, C. Fantozzi, E. Micheloni, N. Pretto, A. Rodà, S. Gasparotto, and G. Salemi. Virtual reconstruction of an ancient greek pan flute. In *Proceedings of the 13th International Sound and Music Computing Conference (SMC'16)*, Hamburg, Germany, September 2016. doi: 10.5281/zenodo.851179.

[10] J. A. Bargas-Avila and K. Hornbæk. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2689–2698. ACM, 2011.

[11] A. Bari, S. Canazza, G. De Poli, and G. A. Mian. Toward a methodology for the restoration of electro-acoustic music. *Journal of New Music Research*, 30(4):365–374, 2001.

[12] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features. *Computer Vision and Image Understanding*, 110(3):346–359, June 2008.

[13] B. Bel and B. Vecchione. Computational musicology. *Computers and the Humanities*, 27 (1):1–5, Jan 1993. ISSN 1572-8412.

[14] A. Berenzweig, B. Logan, D. P. Ellis, and B. Whitman. A large-scale evaluation of acoustic and subjective music-similarity measures. *Computer Music Journal*, 28(2):63–76, 2004.

[15] L. Berio. *Remembering the Future*. The Charles Eliot Norton Lectures, 2006.

[16] N. Bernardini and G. de Poli. The sound and music computing field: Present and future. *Journal of New Music Research*, 36(3):143–148, 2007. doi: 10.1080/ 09298210701862432.

[17] D. M. Berry. Introduction: Understanding the digital humanities. In *Understanding digital humanities*, pages 1–20. Springer, 2012.

[18] A. Bonardi and J. Barthélemy. The preservation, emulation, migration, and virtualization of live electronics for performing arts: An overview of musical and technical issues. *Journal on Computing and Cultural Heritage*, 1(1):6:1–6:16, June 2008. ISSN 1556-4673.

[19] D. Booth, H. Haas, F. McCabe, E. Newcomer, M. Champion, C. Ferris, and D. Orchard. Web services architecture, W3C Working Group Note 11 february 2004. Technical report, W3C, 2004.

[20] G. Boston. *Safeguarding the documentary heritage: a guide to standards, recommended practices and reference literature related to the preservation of documents of all kinds*. UNESCO, 1998.

[21] K. Bradley. *IASA TC-04 Guidelines in the Production and Preservation of Digital Audio Objects: standards, recommended practices, and strategies: 2nd edition*. International Association of Sound and Audio Visual Archives, 2009.

[22] F. Bressan. *The Preservation of Sound Archives: A Computer Science Based Apporach to Quality Control*. PhD thesis, University of Verona, 2013.

[23] F. Bressan and S. Canazza. A systemic approach to the preservation of audio documents: Methodology and software tools. *Journal of Electrical and Computer Engineering*, 2013: 21 pages, 2013.

[24] G. Brock-Nannestad. The objective basis for the production of high quality transfers from pre-1925 sound recordings. In *Proc. Audio Engineering Society 103rd Convention*, pages 26–29, New York, 1997.

[25] T. Brown. *Change by design. How Design Thinking Transforms Organizations and Inspires Innovation.* Harper Collins Business, 2009.

[26] R. Brunelli. *Template matching techniques in computer vision: theory and practice.* John Wiley & Sons, 2009.

[27] D. C. A. Bulterman. Is it time for a moratorium on metadata? *IEEE MultiMedia*, 11(4): 10–17, Oct 2004. ISSN 1070-986X. doi: 10.1109/MMUL.2004.29.

[28] V. Burini, F. Altieri, and S. Canazza. Rilevamenti sperimentali per la conservazione attiva dei documenti sonori su nastro magnetico: individuazione delle curve di equalizzazione. In *Proceedings of the XXI Colloquium of Musical Informatics*, pages 114–121, Cagliari, September 2017.

[29] V. Burini, S. Canazza, and N. Pretto. Archivi di musica elettroacustica: dalla conservazione all'accesso. In *I documenti Sonori. Voce, suono, musica in archivi e raccolte.* Archivi e biblioteche in Piemonte - Direzione Promozione della Cultura, Turismo e Sport della Regione Piemonte, Torino, Italy, 2018.

[30] M. Camras. *Magnetic Recording Handbook.* Van Nostrand Reinhold Co., New York, NY, USA, 1988. ISBN 0442262620.

[31] S. Canazza. The digital curation of ethnic music audio archives: from preservation to restoration. *International Journal of Digital Libraries*, 12(2-3):121–135, 2012.

[32] S. Canazza and A. Vidolin. Preserving electroacoustic music. *Journal of New Music Research*, 30(4):351–363, 2001.

[33] S. Canazza and A. Vidolin. Special issue on Preserving electroacoustic music. *Journal of New Music Research*, 30(4), 2001.

[34] S. Canazza, G. D. Poli, and G. A. Mian. Restoration of audio documents by means of extended kalman filter. *IEEE Trans on Audio Speech and Language Processing*, 18(6): 1107–115, August 2010.

[35] S. Canazza, F. Avanzini, M. M. Novati, and A. Rodà. Active preservation of electrophone musical instruments. the case of the "liettizzatore" of "studio di fonologia musicale" (rai, milano). In S. Zanolla, F. Avanzini, S. Canazza, and A. de Goetzen, editors, *Proc. of Sound and Music Computing Conference*, pages 304–309, 2011.

[36] S. Canazza, C. Fantozzi, and N. Pretto. Accessing tape music documents on mobile devices. *ACM Transactions Multimedia Computing Communications and Applications (TOMM)*, 12(1s):20:1–20:20, October 2015. ISSN 1551-6857. doi: 10.1145/2808200.

[37] R. Caro Repetto, N. Pretto, A. Chaachoo, B. Bozkurt, and X. Serra. An open corpus for the computational research of arab-andalusian music. In *Proceedings of the 5th International Conference on Digital Libraries for Musicology (DLfM '18)*, pages 78–86, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-6522-2. doi: 10.1145/3273024.3273025.

[38] J. M. Carroll. *HCI models, theories, and frameworks: Toward a multidisciplinary science.* Elsevier, 2003.

[39] U. Ceipidor, C. Medaglia, V. Volpi, A. Moroni, S. Sposato, M. Carboni, and A. Caridi. Nfc technology applied to touristic-cultural field: A case study on an italian museum. In *Near Field Communication (NFC), 2013 5th International Workshop on*, pages 1–6, Feb 2013.

[40] A. Celentano and O. Gaggi. Context-aware design of adaptable multimodal documents. *Multimedia Tools and Applications*, 29(1):7–28, Apr 2006. ISSN 1573-7721. doi: 10. 1007/s11042-006-7811-9.

[41] A. Chaachoo. *La música andalusí al-Ála.* Almuzara, Córdoba, 2011.

[42] A. Chaachoo. *La musique hispano-arabe, al-Ala.* L'Harmattan, Paris, 2016.

[43] N. Chomsky. Three models for the description of language. *IRE Transactions on information theory*, 2(3):113–124, 1956.

[44] P. Chordia and A. Rae. Raag recognition using pitch-class and pitch-class dyad distributions. In *ISMIR*, pages 431–436. Citeseer, 2007.

[45] Y. J. Chu and T. H. Liu. On the shortest arborescence of a directed graph. *Science Sinica*, 14:1396–1400, 1965.

[46] E. Civallero. Introducción a las flautas de pan. *Madrid: Civallero, Versión digital*, 13:40, 2014.

[47] E. Cohen. Preservation of audio in folk heritage collections in crisis. *Proceedings of Council on Library and Information Resources*, pages 65–82, 2001.

[48] R. Cohen. *Alan Lomax: Selected Writings, 1934-1997.* Routledge, 2004.

[49] P. R. Cook. *Real sound synthesis for interactive applications.* CRC Press, 2002.

[50] M. Cooley. Human-centered design. *Information design*, pages 59–81, 2000.

[51] P. Copeland. *Manual of analogue sound restoration techniques.* British Library, 2008.

[52] R. Dam and T. Siang. Design thinking: Getting started with empathy. Technical report, Interaction Design Foundation, 2018. URL `https://www.interaction-design.org/literature/article/design-thinking-getting-started-with-empathy`.

[53] K. Damkliang, C. Kaeoaiad, and S. Chehmasong. Ranaad-xek: A prototype design of traditional thai musical instrument application for android tablet pc. In S. Patnaik and X. Li, editors, *Proceedings of International Conference on Soft Computing Techniques and Engineering Application*, volume 250 of *Advances in Intelligent Systems and Computing*, pages 131–140. Springer India, 2014. ISBN 978-81-322-1694-0.

[54] E. D. Daniel. Tape noise in audio recording. *Journal of the Audio Engineering Society*, 20(2):92–99, 1972.

[55] S. W. Davies. Knowledge: The missing element in archiving and restoration? In *Audio Engineering Society Convention 122*. Audio Engineering Society, 2007.

[56] E. De Bono. *New think: The use of lateral thinking in the generation of new ideas*. Avon Books, 1968.

[57] F. de O. Costa, M. A. Oikawa, Z. Dias, S. Goldenstein, and A. R. de Rocha. Image phylogeny forests reconstruction. *IEEE Transaction on Information Forensics and Security*, 9:1533–1546, 2014. ISSN 1556-6013.

[58] M. del Amo. Sistema de transliteración de estudios Árabes contemporáneos. Universidad de Granada. *Miscelánea de Estudios Árabes y Hebraicos. Sección Árabe-Islam*, 51:355–359, 2002.

[59] Z. Dias, S. Goldenstein, and A. Rocha. Exploring heuristic and optimum branching algorithms for image phylogeny. *Journal of Visual Communication and Image Representation*, 24(7):1124 – 1134, 2013. ISSN 1047-3203.

[60] T. Dwyer. *Composing With Tape Recorders: musique concrete for beginners*. Oxford University Press, 1971.

[61] G. Dzhambazov. *Knowledge-based Probabilistic Modeling for Tracking Lyrics in Music Audio Signals*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2017.

[62] J. Edmonds. Optimum branchings. *Journal of research of the National Bureau of Standards - B. Mathematics and Mathematical Physics*, 71B:233–240, 1967.

[63] P. Ehn. *Work-oriented design of computer artifacts*. PhD thesis, Arbetslivscentrum, 1988.

[64] D. A. Eilers. Splicing tapes and their proper application. *Journal of the Audio Engineering Society*, 16(4):472, 474, 476, 1968.

[65] H. Eimert and K. Stockhausen. *Die Reihe 1: Electronic Music*. Theodore Presser Company, 1958.

[66] F. D. Falco and S. Vassos. Museum experience design: A modern storytelling methodology. *The Design Journal*, 20(sup1):S3975–S3983, 2017. doi: 10.1080/14606925.2017. 1352900.

[67] C. Fantozzi, F. Bressan, N. Pretto, and S. Canazza. Tape music archives: from preservation to access. *International Journal on Digital Libraries*, 18(3):233–249, September 2017. ISSN 1432-1300. doi: 10.1007/s00799-017-0208-8.

[68] L. D. Fielder. Pre-and postemphasis techniques as applied to audio recording systems. *Journal of the Audio Engineering Society*, 33(9):649–658, 1985.

[69] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, June 1981.

[70] N. H. Fletcher and T. D. Rossing. *The physics of musical instruments*. New York: Springer-Verlag, 1991.

[71] J. T. Foote. Content-based retrieval of music and audio. In *Multimedia Storage and Archiving Systems II*, volume 3229, pages 138–148. International Society for Optics and Photonics, 1997.

[72] G. Fugazzotto. *Il repertorio musicale dell'emigrazione italiana su dischi a 78 giri: la produzione "made in USA" dei primi trent'anni del Novecento*. PhD thesis, Università degli Studi di Udine, 2009.

[73] S. Godsill and P. Rayner. *Digital Audio Restoration*. Springer-Verlag London, 1998.

[74] S. Godsill, P. Rayner, and O. Cappé. *Digital Audio Restoration*, chapter 4, pages 133–194. Springer US, Boston, MA, 2002. ISBN 978-0-306-47042-4.

[75] N. Goodman. *Languages of art: An approach to a theory of symbols*. Hackett publishing, 1968.

[76] M. Guettat. *La musique arabo-andalouse*. Éditions al-Ouns, Paris, 2000.

[77] S. Hagel. *Ancient Greek music: a new technical history*. Cambridge University Press, 2009.

[78] T. Hall and L. Bannon. Designing ubiquitous computing to enhance children's interaction in museums. In *Proceedings of the 2005 Conference on Interaction Design and Children*, IDC '05, pages 62–69, New York, NY, USA, 2005. ACM. ISBN 1-59593-096-5. doi: 10.1145/1109540.1109549.

[79] R. L. Hess. Tape degradation factors and predicting tape life. In *Audio Engineering Society Convention 121*, Oct 2006.

[80] H.-M. Hua, H. Hayes, H. Hussain, C.-Y. Hsiao, and V. Patel. An active aging app design for older adults. *Gerontechnology*, 13(2):207, 2014.

[81] IEC. Bs en 60094-1:1994 bs 6288-1: 1994 iec 94-1:1981 - magnetic tape sound recording and reproducing systems — part 1: Specification for general conditions and requirements, 1994.

[82] IFLA/UNESCO. *Safeguarding our Documentary Heritage / Conservation préventive du patrimoine documentaire*. UNESCO "Memory of the World" Programme, French Ministry of Culture and Communication, Boston, Massachusetts, USA, 2000.

[83] ITU-R. Bs 1534-1: Method for the subjective assessment of intermediate quality level of coding systems. *International Telecommunication Union*, 2003.

[84] C. R. Johnson, W. A. Sethares, and A. G. Klein. *The envelope of a function outlines its extremes in a smooth manner. Software Receiver Design: Build Your Own Digital Communication System in Five Easy Steps.* Cambridge University Press, 2011.

[85] T. Kell and M. Wanderley. A quantitative review of mappings in musical ios applications. In *Proceedings of the Sound and Music Computing Conference 2013*, pages 473–480, 2013.

[86] L. Kennedy and S.-F. Chang. Internet image archaeology: automatically tracing the manipulation history of photographs on the web. In *ACM International Conference on Multimedia (ACM-MM)*, 2008. ISBN 978-1-60558-303-7.

[87] S. Killingsworth. What makes the hanx writer click? `http://www.newyorker.com/tech/elements/makes-hanx-writer-click`, August 2014.

[88] C. L. Krumhansl and E. J. Kessler. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological review*, 89(4):334, 1982.

[89] O. Lartillot and P. Toiviainen. A matlab toolbox for musical feature extraction from audio. In *International Conference on Digital Audio Effects (DAFx-07)*, pages 237–244, Septempber 2007.

[90] B. Laurel. *Utopian entrepreneur*. MIT Press, Cambridge: MA, 2001.

[91] R. Leydi. *Musica popolare e musica primitiva*. ERI, 1959.

[92] R. Leydi. *I canti poplari italiani: 120 testi e musiche*. Mondadori, 1973.

[93] R. Leydi. L'altra musica. *Firenze/Milano*, 1991.

[94] G. D. Lord, B. Lord, A. Hossaini, and N. Blankenberg. *Manual of Digital Museum Planning*. Rowman & Littlefield Publishers, Inc., 2017. ISBN 144227896X, 9781442278967.

[95] M. Lykke and C. Jantzen. User experience dimensions: A systematic approach to experiential qualities for evaluating information interaction in museums. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, pages 81–90. ACM, 2016.

[96] S. MacLeod, J. Dodd, and T. Duncan. New museum design cultures: harnessing the potential of design and 'design thinking' in museums. *Museum Management and Curatorship*, 30(4):314–341, 2015. doi: 10.1080/09647775.2015.1042513.

[97] P. Magaudda. When materiality 'bites back': Digital music consumption practices in the age of dematerialization. *Journal of Consumer Culture*, 11(1):15–36, 2011.

[98] J. C. Mallinson. Tutorial review of magnetic recording. *Proceedings of the IEEE*, 64(2): 196–208, Feb 1976. ISSN 0018-9219.

[99] A. Melloni, P. Bestagini, S. Milani, M. Tagliasacchi, A. Rocha, and S. Tubaro. Image phylogeny through dissimilarity metrics fusion. In *European Workshop on Visual Information Processing (EUVIP)*, 2014.

[100] A. Menegazzi, M. Cesaretto, E. M. Ciampini, and P. Zanovello. La scatola misteriosa. In P. Zanovello and E. M. Ciampini, editors, *Egitto in Veneto*, pages 91–104, 2013.

[101] E. Micheloni, N. Pretto, F. Avanzini, S. Canazza, and A. Rodà. Installazioni interattive per la valorizzazione di strumenti musicali antichi: il flauto di pan del museo di scienze archeologiche e d'arte dell'università degli studi di padova. In *Extending interactivity. Atti del XXI CIM - Colloquio di Informatica Musicale (CIM'16)*, pages 203–208, Cagliari, Italy, September 2016. AIMI - Associazione Informatica Musicale Italiana.

[102] E. Micheloni, N. Pretto, and S. Canazza. A step toward ai tools for quality control and musicological analysis of digitized analogue recordings: recognition of audio tape equalizations. In *Proceedings of the 11th International Workshop on Artificial Intelligence for Cultural Heritage (AI*CH'17)*, pages 17–24, Bari, Italy, November 2017. URL `www.scopus.com/inward/record.uri?eid=2-s2.0-85041462080&partnerID=40&md5=0fabe08bda617c677dfa785145504825`.

[103] S. Milani, M. Fontana, P. Bestagini, and S. Tubaro. Phylogenetic analysis of near-duplicate images using processing age metrics. In *IEEE ICASSP*, 2016.

[104] S. Milani, P. Bestagini, and S. Tubaro. Video Phylogeny Tree Reconstruction Using Aging Measures. In *Proc. of the 2017 European Signal Processing Conference (EUSIPCO 2017)*, Kos, Greece, Aug. 28 – Sep. 2 2017.

[105] M. Miliano, editor. *The IASA cataloguing rules (IASA 1999)*. International Association of Sound and Audiovisual Archives IASA Editorial Group, 1999.

[106] NAB. Magnetic tape recording and reproducing (reel-to-reel), 1965.

[107] K. Ng, R. Boyle, and D. Cooper. Automatic detection of tonality using note distribution. *Journal of New Music Research*, 25(4):369–381, 1996.

[108] M. M. Novati and J. Dack. *The Studio Di Fonologia: a musical journey 1954-1983, update 2008-2012*. Ricordi, 2012.

[109] M. Nucci, M. Tagliasacchi, and S. Tubaro. A phylogenetic analysis of near-duplicate audio tracks. In *2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP)*, pages 099–104, Sept 2013.

[110] A. R. Olesen. For the sake of technology? The role of technology views in funding and designing digital museum communication. *Museum Management and Curatorship*, 31(3): 283–298, 2016. doi: 10.1080/09647775.2016.1163643.

[111] C. Pendharkar, P. Bäck, and L. Wyse. Adventures in scheduling, buffers and parameters: Porting a dynamic audio engine to web audio. In *Proceedings of the Web Audio Conference, Paris, France*, 2015.

[112] C. Poché. *La música arábigo-andaluza*. Akal, Móstoles, 2005.

[113] A. Porter, M. Sordo, and X. Serra. Dunya: A system for browsing audio music collections exploiting cultural context. In *14th International Society for Music Information Retrieval Conference (ISMIR 2013)*, pages 101–106, Curitiba, Brazil, 2013.

[114] H. Pousseur. *La musica elettronica: testi*. Feltrinelli, 1976.

[115] H. Pousseur. *Ecrits théoriques, 1954–1967*. Sprimont: Editions Pierre Mardaga, 2004.

[116] N. Pretto and S. Canazza. Rewind: Simulazione di un'esperienza d'ascolto storicamente fedele di dischi fonografici digitalizzati. In *Proceedings of the XX Colloquium of Musical Informatics, (CIM'14)*, pages 29–34, Roma, Italy, October 2014.

[117] N. Pretto, C. Fantozzi, E. Micheloni, V. Burini, and S. Canazza. Computing methodologies supporting preservation of electroacoustic music from analog magnetic tape. *Computer Music Journal*, (Accepted in November 2018).

[118] N. Pretto, B. Bozkurt, R. Caro Repetto, and X. Serra. Nawba recognition for arab-andalusian music using templates from music scores. In *15th Sound and Music Computing Conference (SMC'18)*, pages 405–410, Limassol, Cyprus, July 2018. doi: 10.5281/zenodo.1257388.

[119] C. Roberts, G. Wakefield, M. Wright, and J. Kuchera-Morin. Designing musical instruments for the browser. *Computer Music Journal*, 39(1):27–40, March 2015. ISSN 0148-9267.

[120] F. Rumsey. Challenges in archiving and restoration. *Journal of the Audio Engineering Society*, 64(1/2):94–97, 2016.

[121] S. S. *Computational Analysis of Audio Recordings and Music Scores for the Description and Discovery of Ottoman-Turkish Makam Music*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2016.

[122] D. Saffer. *Designing for Interaction: Creating Smart Applications and Clever Devices*. Safari Books Online. New Riders, 2007. ISBN 9780321432063. URL https://books.google.es/books?id=D9yNQgAACAAJ.

[123] F. Sallis, V. Bertolani, J. Burle, and L. Zattra. *Live-Electronic Music. Composition, Performance and Study*. Routledge, 2017.

[124] E. B.-N. Sanders. Postdesign and participatory culture. In *Proceedings of Useful and Critical: The Position of Research in Design*, Helsinki, 1999. University of Art and Design.

[125] E. B.-N. Sanders and P. J. Stappers. Co-creation and the new landscapes of design. *Co-design*, 4(1):5–18, 2008.

[126] D. Schüller. The ethics of preservation, restoration, and re-issues of historical sound recordings. *Journal of the audio engineering society*, 39(12):1014–1017, 1991.

[127] D. Schüller. Preserving the facts for the future: Principles and practices for the transfer of analog audio documents into the digital domain. *Journal of Audio Engineering Society*, 49(7-8):618–621, 2001.

[128] X. Serra. Creating research corpora for the computational study of music: the case of the compmusic project. In *Audio Engineering Society Conference: 53rd International Conference: Semantic Audio*, pages 1–9, London, United Kingdom, Jan 2014.

[129] X. Serra. The computational study of a musical culture through its digital traces. *Acta Musicologica*, 89(1):24–44, 2017. URL `http://hdl.handle.net/10230/32302`.

[130] M. Sorce Keller. En souvenir de roberto leydi (1928-2003). les origines et l'évolution de l'ethnomusicologie en italie. *Cahiers d'ethnomusicologie. Anciennement Cahiers de musiques traditionnelles*, (17):297–314, 2004.

[131] M. Sordo, A. Chaachoo, and X. Serra. Creating corpora for computational research in Arab-Andalusian music. In *1st International Workshop on Digital Libraries for Musicology (DLfM 2014)*, pages 1–3, London, United Kingdom, 2014. ACM.

[132] U. D. Staso, D. Magliocchetti, and R. D. Amicis. Smart-islands: Enhancing user experience for mediterranean islands for tourism support. In A. Marcus, editor, *Proceedings of DUXU 2014*, volume 8518 of *Lecture Notes in Computer Science*, pages 223–233. Springer, Heraklion, Crete, Greece, June 2014.

[133] M. B. Stogner. The media-enhanced museum experience: Debating the use of media technology in cultural exhibitions. *Curator: The Museum Journal*, 52(4):385–397, 2009. ISSN 2151-6952. doi: 10.1111/j.2151-6952.2009.tb00360.x.

[134] W. Storm. The establishment of international re-recording standards. *Phonographic Bulletin*, 27:5–12, 1980.

[135] *Studer A810 - Operating and service instruction*. Studer, 1982.

[136] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, June 2015.

[137] J. Szwed. *Alan Lomax: The man who recorded the world*. Penguin, 2010.

[138] D. Temperley and E. W. Marvin. Pitch-class distribution and the identification of key. *Music Perception: An Interdisciplinary Journal*, 25(3):193–212, 2008.

[139] S. Timpanaro. *The Genesis of Lachmann's Method*. University of Chicago Press, 2005.

[140] A. Uitdenbogerd and J. Zobel. Melodic matching techniques for large music databases. In *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, pages 57–66. ACM, 1999.

[141] UNESCO. *Convention for the safeguarding of the intangible cultural heritage*. UNESCO, Paris, 2003.

[142] V. Valimaki and J. D. Reiss. All about audio equalization: Solutions and frontiers. *Applied Sciences*, 6(5), 2016. ISSN 2076-3417.

[143] S. Verde, S. Milani, P. Bestagini, and S. Tubaro. Audio phylogenetic analysis using geometric transforms. In *2017 IEEE International Workshop on Information Forensics and Security (WIFS)*, Dec. 2017.

[144] S. Verde, N. Pretto, S. Milani, and S. Canazza. Stay true to the sound of history: Philology, phylogenetics and information engineering in musicology. *Applied Sciences*, 8(2), 2018. ISSN 2076-3417. doi: 10.3390/app8020226.

[145] L. Wein. Visual recognition in museum guide apps: Do visitors want it? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 635–638, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2473-1.

[146] D. Williams, A. Pooransingh, and J. Saitoo. Efficient music identification using orb descriptors of the spectrogram image. *EURASIP Journal on Audio, Speech, and Music Processing*, 2017(1):17, Jul 2017. ISSN 1687-4722.

[147] L. Wyse and S. Subramanian. The viability of the web browser as a computer music platform. *Computer Music Journal*, 37(4):10–23, Dec 2013. ISSN 0148-9267.

[148] M. Zahedi. Designer as mediator: An innovative model for designing interactive devices within the user-centered approach. In *Eighth IEEE International Conference on Advanced Learning Technologies (ICALT'08)*, pages 344–345. IEEE, 2008.

[149] M. Zanoni, S. Lusardi, P. Bestagini, A. Canclini, A. Sarti, and S. Tubaro. Robust music identification approach based on local spectrogram image descriptors. In *Proc. of 142th AES Convention*, Berlin, Germany, May 2017. paper no. 9763.

[150] L. Zattra. The assembling of stria by john chowning: A philological investigation. *Computer Music Journal*, 31(3):38–64, 2007.

[151] L. Zattra, G. De Poli, and A. Vidolin. Yesterday sounds tomorrow. preservation at csc. *Journal of New Music Research*, 30(4):407–412, 2001.

[152] S. Zhang, R. Caro Repetto, and X. Serra. Understanding the expressive functions of jingju metrical patterns through lyrics text mining. In *18th International Society for Music Information Retrieval Conference*, pages 397–403, Suzhou, China, Oct 2017. URL `http://hdl.handle.net/10230/32652`.