IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

From Tape to Code: An international AI-based standard for audio cultural heritage preservation

Don't play that song for me (if it's not preserved with ARP!)

MARINA BOSI[1], SERGIO CANAZZA[2], NICCOLÒ PRETTO[3], ALESSANDRO RUSSO[2], and MATTEO SPANIO [2]

[1]Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, 660 Lomita Ct, Stanford, CA 94305 USA
[2]Centro di Sonologia Computazionale (CSC), Department of Information Engineering (DEI), University of Padova, Via G. Gradenigo 6/b, Padua, 35131 Italy
[3]Media Interaction Lab, Faculty of Engineering, Free University of Bozen-Bolzano, NOI Techpark, Via A. Volta 13, Bozen, 39100 Italy

Corresponding author: S. Canazza (e-mail: sergio.canazza@unipd.it).

ABSTRACT This article describes a novel technology for preserving audio documents archived on open-reel magnetic tapes forming the core of the Audio Recording Preservation (ARP) international standard. ARP is part of the Moving Picture, Audio, and Data Coding by Artificial Intelligence (MPAI) Context-based Audio Enhancement (CAE) standard, adopted by the IEEE Standard Association as IEEE 3302-2022 in December 2022. Leveraging automated Artificial Intelligence (AI) tools, ARP analyzes and extracts relevant information from digitized audio and video files of the tape's corresponding digital Preservation Copy. This process includes identifying speed variations and surface irregularities on the tape, automatically rectifying errors to generate a restored Access Copy. By utilizing the ARP standard, archives gain a potent tool for expediting and optimizing the description of the preservation conditions of the tape, as well as automatically correcting any errors that may have occurred during the digitization process. This technology offers an efficient solution for managing both small and large collections of digitized analog items, marking a substantial advancement in the preservation of audio documents.

INDEX TERMS artificial intelligence, audio documents preservation, audio restoration, IEEE standard, musicological analysis, MPAI standard

I. INTRODUCTION

> In the summer of 1937, Bird [Charlie Parker's nickname, one of the most important jazz musicians of the twentieth century - Ed. note] underwent a radical change musically. He got a job with a little band led by a singer... they played at country resorts in the mountains. Charlie took with him all the Count Basie records with Lester Young solos on them and learned Lester cold, note for note... when he came back, only two or three months later, the difference was unbelievable. (Gene Ramey [1])

THe legendary Charlie Parker stands as a compelling illustration of how musical documents can shape the course of history. In the early stages of his career, Bird immersed himself in the records of Lester Young—a well-documented instance of a virtuoso jazz musician learning from another. This dynamic exchange gave rise to a new realm of musical improvisation, building upon the foundation laid by previous masters. Such creative evolution would have been inconceivable without easy access to the records of his predecessor. While the preservation of these cultural documents assumes paramount importance and continues to provide incredible opportunities for future generations, it also poses ongoing challenges and rewards, particularly with the advent of new technologies rooted in AI.

When preserving cultural audio heritage, it is fundamental to minimize loss of information. This is particularly significant when dealing with genres like Afro-American jazz music, where a traditional score might not exist. Additionally, in cases such as Tape Music, where the magnetic tape itself is an integral part of the artistic work, careful preservation is essential to safeguard the complete artistic experience.

Unlike recordings of live musicians, Tape Music is not

FIGURE 1. Example of markings on a tape splice.

captured on stage or in the studio for later storage and reproduction. Instead, it is composed directly with the assistance of electronic valves, transistors, and similar devices. Tape Music "exists" exclusively on magnetic tapes and can be reproduced and experienced through loudspeakers. The viability of these techniques in music composition emerged with the introduction of magnetic tape sound recording technologies. These advancements enabled direct human manipulation and (acoustic-)electromagnetic treatment of the recording medium. As a result, Tape Music captured the interest of prominent experimental and avant-garde creative minds in the mid-twentieth century. Notable figures such as Edgard Varèse (1883–1965), Olivier Messiaen (1908–1992), John Cage (1912–1992), Iannis Xenakis (1922–2001), Luigi Nono (1924–1990), Luciano Berio (1925–2003), Pierre Boulez (1925–2016), and Karlheinz Stockhausen (1928–2007) were drawn to explore its possibilities.

In this context, the primary challenges arise from the relatively short life expectancy of this medium (less than 20 years), in contrast to the longevity of conventional tangible cultural heritage, which can endure for centuries or even millennia. This situation calls for a transition to re-recording these documents in digital form, ensuring their preservation over time. Relying solely on audio copies, however, is insufficient for preservation. Composers actively engaged with the tape, employing techniques such as cutting and pasting, adding annotations directly onto the medium (as illustrated in Fig. 1). Some clues are essential for live performances of the piece, while others, though not directly impacting performance, hold significance from a philological standpoint. Often, composers did not furnish a traditional score; therefore, the tape itself becomes the artwork—the culmination of the creative process. Preserving the tape in its entirety is crucial to safeguarding the essence of the artistic creation.

The integration of electronic and information technology into art has presented fresh challenges for archives and the preservation of cultural heritage. While technology serves as a catalyst for innovative forms of artistic creation, it also contributes to the accelerated deterioration and deprecation of formats, thereby reducing the lifespan and accessibility of new artworks.

Critical issues in this context include the compounded sheer volume of material yet to be digitized and the variety of adopted formats. The risks are twofold: firstly, the lack of expertise in digitization may result in the loss of information, and secondly, the limited storage space and bandwidth available pose significant hurdles in the challenge of preserving these archives for posterity. Data analysis, which may occur years after digitization, may highlight error inconsistencies in audio documents that are no longer easily accessible in the original format. Overall, digitization and data analysis represent a significant investment, requiring considerable resources in terms of time, money, and technical expertise. Naive implementations may jeopardize the proper preservation and accessibility of cultural heritage, making it unattainable.

These issues are addressed by the novel technology outlined in the Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI) international standard on Audio Recording Preservation (ARP), later adopted as IEEE 3302-2022. Drawing extensively from contributions of the Centro di Sonologia Computazionale (CSC) at the University of Padova [2], [3], and leveraging considerable experience in music production, the ARP approach revolves around a well-defined scientific methodology anchored on two essential pillars. Firstly, it adopts a multidisciplinary approach that integrates perspectives from engineers, musicians, musicologists, composers, and archivists. Secondly, it upholds a profound commitment to philological accuracy in the development of digital tools. This encompasses the inclusion of metadata and ancillary information deemed crucial for the comprehensive completion of preservation copies [4]. The datasets gathered at CSC were assembled from over 3000 documents digitized through numerous preservation projects [5]. Notably, some of the most representative restored and digitized collections include those from Luciano Berio's archive (Paul Sacher Stiftung, featuring tape music and electronic music), the Luigi Nono Archive of Venice (tape music, electronic music), the Historical Archive of the Teatro Regio of Parma (encompassing opera, Western classical music, and pop/rock), the Tullia Magrini Archive (focused on ethnomusic), the Historical Archive of the Maggio Musicale Fiorentino (covering Opera and Western classical music), and the Fondazione Giorgio Cini of Venice (comprising speech and oral sources).

The structure of this manuscript is as follows. Section II briefly examines existing guidelines, relevant literature, and solutions for audio document preservation, along with the application of AI in music. Section III delves into the preservation methodology forming the core of the ARP standard. Section IV presents an in-depth overview of the foundational infrastructure and novel technology underpinning ARP, while Section V summarizes the performance results of the various technology components adopted, concluding with final remarks.

II. STATE OF THE ART

To ensure the preservation of audio documents, it's imperative to establish internationally shared guidelines that set

preservation standards. These guidelines should encompass a regulatory framework that addresses the various stages of the preservation process, such as digitization, archiving, and long-term preservation of audio documents. International organizations, such as the International Association of Sound and Audiovisual Archives (IASA) [6] and the International Federation of Library Associations and Institutions (IFLA) [7] have contributed to the definition of protocols aimed at guaranteeing the quality and sustainability of preservation practices for diverse physical media. Such guidelines, however, often inadequately describe the organization of digitized files, primarily focusing on the correct practices for preserving and managing analog documents. The organization of digitized files, encompassing metadata and storage formats, demands meticulous planning to guarantee the long-term accessibility and integrity of the contents. Hence, shared guidelines must comprehensively embrace the evolving digital landscape, addressing the challenges and best practices pertinent to preserving audiovisual documents in digital formats. To tackle these challenges, the CSC has proposed to MPAI a preservation methodology for audio documents based on [8], elaborated in detail in Section III.

In recent years, numerous archives and private institutions have embarked on extensive digitization projects. These endeavors, however, often encounter the challenge of digitizing a vast quantity of audiovisual documents within a relatively short time frame, which can easily lead to errors during the digitization process. The pressure to meet deadlines while handling such a large volume of materials may result in oversights or mistakes, ultimately compromising the quality and accuracy of the digitized records. Issues such as incomplete signal transfers, mislabeled files, or inadequate preservation of metadata may arise from these digitization efforts. Therefore, archives must allocate sufficient time and resources to minimize the risk of errors when digitizing analog materials. In this regard, AI emerges as an invaluable tool for enhancing efficiency and accuracy in the digitization process.

The integration of AI into the realm of music has begun to revolutionize how artists, composers, and producers approach music composition [9], creation, and production. AI's capacity to analyze extensive musical datasets, discern patterns, and identify genres [10] enables it to generate novel sounds. Artists can harness AI for inspiration, crafting innovative melodies, and exploring unique sonic landscapes. Moreover, in music production, AI can optimize processes such as mixing and mastering, enhancing overall sound quality. Some AI-based tools even facilitate automatic composition [11] of personalized musical accompaniments or real-time adaptation of music to listeners' emotions. Alongside these creative opportunities, however, concerns regarding ethics and artistic integrity emerge regarding AI's potential to supplant the human element in music creation and the preservation of artistic authenticity. As of now, the use of AI in audio preservation has remained relatively limited, primarily focusing on speech restoration [12], [13], quality assessment of digitized audio [14], and, only very recently, historical recording restoration issues [15]. Although AI has found significant applications in music creation and production, its adoption in the conservation of historical audio recordings and management of music archives is still in its very early stages. Nowadays, archives are facing a digital transformation and they must make use of automation to manage data, especially in the form of AI [16].

The primary challenge lies in the intricacy and sensitivity of audio preservation, necessitating a meticulous and reverent approach to maintain the quality and authenticity of recordings over time. The research presented in this paper concentrates on investigating novel applications of AI in the restoration and conservation of audio recordings.

III. PRESERVATION METHODOLOGY

The preservation methodology for audio documents, originally developed by the CSC and officially accepted and implemented as part of the ARP standard, is illustrated in Fig. 2. The initial step of the methodology involves photographing each audio document along with its corresponding box to document its preservation status. This information is crucial, as composers frequently made annotations on the boxes, covering not only details about the recording contents but also the adopted channel configuration, equalization curve, and recording speed. While the potential for misalignment between the content and what is reported on the boxes exists, it still serves as a valuable guideline. Visual inspection and pre-reading play an important role in diagnosing evident mechanical issues or identifying chemical/physical syndromes that may impact the tape. These steps provide essential insights before proceeding with the restoration process. Overall, the optimization of the carrier includes fixing old splices through the original tape, applying leader tape at the beginning, cleaning the surface to remove mold and dust, and implementing thermal treatment to address the Sticky-Shed/Soft Binder Syndrome [17], [18].

The second step of the methodology concerns the A/D conversion. To digitize the audio content, it is fundamental to analyze and set the recording formats, digital parameters, and playback configuration correctly. Monitoring the entire A/D process is essential for preventing errors, such as the misinterpretation of channel configuration and recording speed. The digitization process is executed with high-quality converters and fully operational analog devices. Video documentation of the tape is also included to track any irregularities that may be present on its surface.

The final step of the preservation methodology involves data processing and the creation of Preservation and Access Copies. The Preservation Copy comprises a high-quality digital audio file with audio stored at a minimum of 24 bits precision and a sampling rate of 96 kHz, without any restoration or filters applied. In the case of multi-channel recordings, a separate audio file is provided for each channel. Multiple acquisitions are conducted when a tape is recorded at different speeds, resulting in separate audio files. In addition to digital audio files, the Preservation Copy incorporates photographic

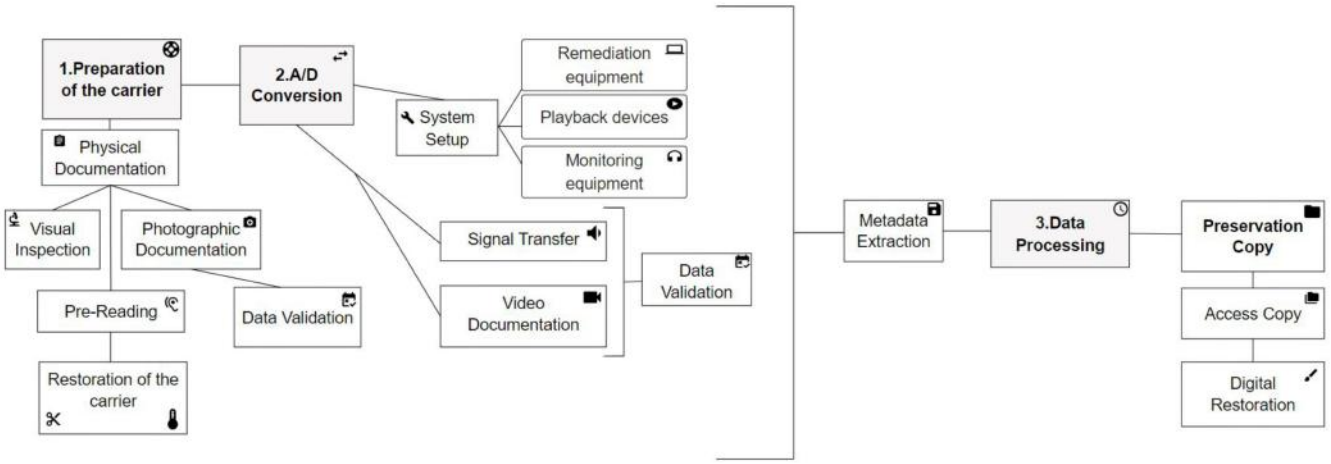

FIGURE 2. Diagram illustrating the preservation methodology.

and video documentation, checksums, and scanned images of any accompanying documentation that may have been with the original item. Metadata gathering plays a central role in this process. Data regarding the original document, including brand, reel diameter, channel configuration, recording speed, etc., is stored in a dedicated database and summarized in a .pdf file, which is also included in the Preservation Copy. The Access Copy is typically provided in a compressed format, such as MPEG AAC [19], [20], to enhance portability and can be digitally restored if necessary.

One of the most common challenges in digitizing analog audio tapes is applying the correct equalization (EQ) curve. EQ curves were employed during recording as pre-emphasis to extend the dynamic range and enhance the Signal-to-Noise Ratio (SNR). During playback, inverse post-emphasis curves were applied to restore the original frequency response. Identifying the correct EQ curve is a significant challenge, particularly when dealing with tapes recorded in the early days of sound recording when there were no shared standards. In certain instances, different record labels and/or even individual technicians might have chosen to apply customized EQ curves to improve sound quality or tailor it to the technical characteristics of the equipment used at that time. The introduction of standard EQ curves such as IEC1 [21] (formerly known as CCIR) and IEC2 [22] (formerly known as NAB) has streamlined this process, yet it does not entirely resolve the issue. The digitization process remains complex, as it necessitates identifying and correcting the EQ curves to ensure an accurate and faithful reproduction of the original sound.

IV. CAE-ARP

The ARP technology is part of the MPAI-CAE international standard (aka IEEE 3302-2022[1]). MPAI/IEEE-CAE's pioneering specifications extend across a wide array of applications, including entertainment, communication, teleconferencing, gaming, post-production, preservation and restoration [23]. MPAI/IEEE-CAE encompasses four distinct use cases tailored to enhance the user's audio experience across various contexts, spanning different settings such as the home, car, on-the-go, and studio. The four use cases specified in the CAE standard are: 1) Emotion Enhanced Speech (EES); 2) Audio Recording Preservation (ARP); 3) Speech Restoration System (SRS); 4) Enhanced Audioconference (EAE). These examples highlight the versatility and comprehensive scope of MPAI/IEEE-CAE's innovative specifications, illustrating their adaptability to various contexts and their ability to address a broad spectrum of audio-related needs [23].

The foundational infrastructure enabling the implementation of MPAI-CAE is the MPAI AI Framework (AIF), specified in the MPAI-AIF/IEEE 3301-2022 standard[2]. This provides the operational backbone for executing AI Workflows (AIW), which are constructed from fundamental processing elements known as AI Modules (AIM). MPAI-CAE normatively defines the semantics and syntax of input and output data, the functions of the AIW and AIMs, as well as the connections between AIMs within an AIW. Interoperability is ensured by the ability to substitute an AIW or AIM implementation with a functionally equivalent one while maintaining correct input/output formats. MPAI-CAE's objective is to leverage this embedded structure to enhance user experiences in audio-related applications.

The CAE-ARP technology stands as a groundbreaking advancement in the accurate preservation of information found in open-reel audio tapes. Through this process, not only long-term preservation but also precise playback of the digitized recording is ensured, with the capability for

[1]https://standards.ieee.org/ieee/3302/11006/ Last accessed March 19, 2024

[2]https://standards.ieee.org/ieee/3301/11096/ Last accessed March 19, 2024

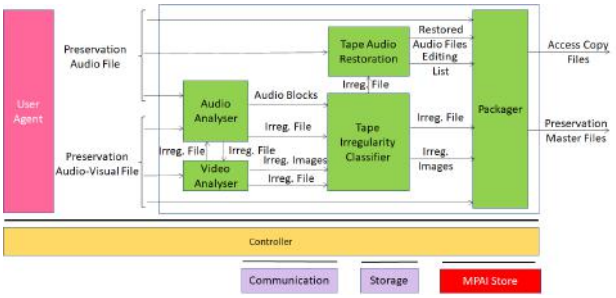

FIGURE 3. MPAI-CAE ARP AI Workflow.

restoration if needed. CAE-ARP leverages automated AI processes to extract crucial information from digitized audio files, facilitating the creation of preservation and access copies. Operating within the framework of the CAE-ARP standard, archives can efficiently manage the wealth of information stored on tapes and their associated metadata. This standardized approach enables the automated preparation of content for immediate storage and/or utilization, streamlining the archival process and enhancing accessibility.

The ARP AIW and its various components are illustrated in Fig. 3. The architecture of the ARP standard comprises five AIMs designed to target and process distinct digital inputs [24]. These include the *Audio Analyzer*, *Video Analyzer*, *Tape Irregularity Classifier*, *Tape Audio Restoration*, and *Packager*. Each AIM plays a specific role in the overall processing and enhancement of audio content, contributing to the comprehensive capabilities of the ARP technology. Preserving audio assets recorded on analog media holds significant importance, considering the valuable information embedded in the magnetic tape of an open reel. In addition to the audio signal, this information may include annotations by the composer or technicians, multiple splices, and various irregularities like carrier corruptions, different-colored tapes, or diverse chemical compositions. The primary ARP objectives are long-term preservation and the creation of an access copy, which is restored if necessary, to facilitate accessibility and correct playback of the digitized recording. The ARP process takes as input the Preservation Audio File, which is generated through the digitization of the analog audio signal recorded on an open-reel tape with 24 bits per audio sample and a sampling rate of 96 kHz. Furthermore, an essential input to the ARP is the Preservation Audio-Visual File, which amalgamates a video file generated by a camera positioned at the playback head of the open-reel tape machine, see Fig. 4, with the audio content digitized at low resolution and synchronized with the video file. This comprehensive input contributes to the preservation process, ensuring that both audio and visual elements are accurately captured and maintained. The first AIMs in the ARP AIW (see Fig. 3), the Audio and Video Analyzers, analyze the audio/video signals in order to detect irregularities (such as Splice, Brands on tape, Start of tape, Ends of tape, Damaged tape, Dirt, Marks, Shadows, Wow and flutter, Play, pause and stop, Speed standard variation, Equalization standard variation, Signal backward) and create an Irregularity File and associated Audio and Image Files. These files feed into the Tape Irregularity Classifier AIM which classifies and selects the ones considered relevant. If the selected Irregularity was detected by the Video Analyzer, in addition to the selected Irregularity File, the corresponding Irregularity Images are also sent to the Packager AIM. The Tape Audio Restoration AIM uses the Irregularity File to identify and restore portions of the Preservation Audio File. It corrects speed, equalization and reading backwards errors in the Preservation Audio File and sends the Restored Audio Files and an Editing List to the Packager AIM. Finally the Packager AIM collects the Preservation Audio Files, Restored Audio Files, the Editing List, the Irregularity File and corresponding Irregularity Images if detected by the Video Analyzer, and the Preservation Audio-Visual File and it produces the Preservation Master Files (which contain the Preservation Audio File, the Preservation Audio-Visual File where the audio has been replaced with the reduced resolution audio of the Preservation Audio File fully synchronised with the video, the set of Irregularity Images and the Irregularity File) and Access Copy Files (which contain the Restored Audio Files, the Editing List, the set of Irregularity Images and the Irregularity File). In the following sections, we will provide an in-depth description of the key technologies that are employed in implementing the various MPAI-CAE ARP AIMs. Before diving into the technical details, we will first offer an overview of MPAI, an international, unaffiliated non-profit organization committed to establishing standards for data coding based on AI.

A. MPAI

Established in September 2020 in Geneva, MPAI is an international standards organization committed to advancing the efficient utilization of data. Its mission involves developing technical specifications across diverse fields [25], encompassing Audio, Video, Neural Network Watermarking, Human-Machine Interaction, Avatars, Metaverse, Real and Virtual Environment Performance, Online Gaming, Financial Data, and Health. MPAI operates at the forefront of innovation, incorporating new technologies such as AI to shape standards that address the evolving landscape of data-related applications.

In its first three years of existence, MPAI has successfully developed and released 9 standards, all of which are publicly accessible on their website[3]. Notably, 5 of these standards have been officially adopted by the IEEE Standards Association (IEEE SA), showcasing MPAI's influence and contribution to the broader standards community. Additionally, MPAI continues to work on and has more standards in the pipeline, underscoring its ongoing commitment to advancing technological standards in various domains.

Apart from the technical specifications, MPAI has developed 3 reference software implementations, which are

[3]https://mpai.community/standards/ Last accessed March 19, 2024

publicly accessible as open source. Furthermore, MPAI has released 2 conformance testing specifications publicly, offering valuable resources for evaluating adherence to standards. Lastly, MPAI has introduced 1 performance assessment specification, publicly available, that assesses factors such as robustness, replicability, reliability, and fairness, providing insights into the effectiveness and dependability of the implemented standard.

For MPAI-CAE ARP, the technical specifications as well as the conformance testing specifications and reference software are available through the MPAI website[4]. CAE ARP stands out as one of MPAI's most successful technologies, being recognized twice (in 2023 and in 2024) by the prestigious "Neurons Awards Creativity AI Trophy" at the World Artificial Intelligence Cannes Festival (WAICF[5]), the world's largest artificial intelligence event[6].

In the following sections, a comprehensive technical description of the ARP AIMs is presented.

B. AIW ARCHITECTURE IMPLEMENTATION

The current infrastructure of ARP is implemented through a set of docker containers that interact via the Remote Procedure Call (RPC) protocol [26] (using gRPC implementation) and share a volume where to store the data. Each docker container hosts a server with a module implementation and exposes an API. This entire setup is managed by a client that sends organized requests to the services and processes their responses. More specifically, a common interface for all ARP AIMs has been defined via Protocol Buffer (aka Protobuf)[7], which exposes a main method, called work, for starting data processing in each module and receiving responses based on their current state. The Protobuf interface is currently implemented in Python.

The code for this infrastructure implementation, along with its documentation, is also available on Gitlab[8]

C. AUDIO ANALYZER AND VIDEO ANALYZER

The first two AIMs within the ARP AIW, as illustrated in Fig. 3, namely the Audio and Video Analyzers, are specifically designed to detect tape irregularities and accurately determine the exact moment at which these irregularities occur. The input to the ARP standard comprises two distinct files: a Preservation Audio File (PAF) obtained through the high-quality digitization of the analog audio, encompassing music, soundscape, or speech, recorded on the magnetic tape; and a Preservation Audio-Visual File (PAVF) created by a camera focused on the reading head of the magnetic tape machine (see Fig. 4). Together, these files contribute to the comprehensive preservation of both audio and visual aspects of the magnetic tape content. The PAF plays a crucial role in identifying any errors in the application of EQ curves, tape speed, and reverse audio [27] and it is then processed to be both restored and archived unaltered for philological purposes. In addition, the PAVF proves valuable for managing metadata associated with the carrier and providing additional information related to the context of the recording. This dual-input approach enhances the accuracy and thoroughness of the preservation process within the ARP standard.

1) Video Analyzer

A fundamental aspect of the Video Analyzer module is the precise identification of Regions of Interest (ROIs). Preliminary studies explored the application of background subtraction algorithms, utilizing prior information to segregate new elements from recurring ones. This approach, however, exhibited limitations, primarily manifesting as false positives due to variations in brightness, reel movement, and undesired artifacts [28]. In response to these challenges, a paradigm shift towards a scene framing-oriented approach was undertaken. It was observed that anomalies consistently exhibited a lack of vertical movement, appearing as small clusters of points within the frame. The strategic decision to focus on stationary elements, notably the capstan and reading head (see Fig. 4), was made to serve as reference points for automatically identifying pixel regions vulnerable to irregularities [29]. The reading head, a pivotal component in tape recorders, proved to be a salient reference point due to its inherent stationary nature during playback. Coupled with the capstan, these components facilitated the establishment of reliable stationary elements.

A thorough analysis of the central frame of the video associated with the tape, which is assumed to be indicative of a standard scenario, is carried out. After extracting the image (with deinterlacing applied in the case of older PAL videos), the positioning of ROIs is determined by seeking correspondence within grayscale capstan and reading head reference images. The element individuation is achieved through the utilization of the well-established Generalized Hough Transform [30] and SURF [31] algorithms, while the transformation from RGB space to grayscale is dependant to OpenCV [32], the library used in the actual implementation to perform image processing, which defines the conversion rule as follow:

$$\text{RGB[A] to Gray: } Y \leftarrow 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B$$

[4]https://mpai.community/standards/mpai-cae/ Last accessed March 19, 2024

[5]https://www.worldaicannes.com/en Last accessed March 19, 2024

[6]https://web.archive.org/web/20231210130015/https://www.worldaicannes.com/en/cannes-neurons links to the 2023 results and https://www.worldaicannes.com/en/cannes-neurons links to the 2024 results. Last accessed March 19, 2024

[7]https://protobuf.dev/ Last accessed March 19, 2024

[8]The CAE-ARP reference software can be found at the following link: https://experts.mpai.community/software/mpai-private/mpai-cae/arp/arp-workflow Last accessed on March 22, 2024. At the time of writing, accessing the official MPAI GitLab instance requires permission from the system administrator. Hence, to ensure accessibility for the reviewers of this article, a corresponding repository has been established on the University of Padua server. This repository encompasses the infrastructure implementation of the standard, providing the interface for the AIMs. This repository is publicly accessible via the link: https://gitlab.dei.unipd.it/csc-research/arp-aiw. Last accessed on March 22, 2024.

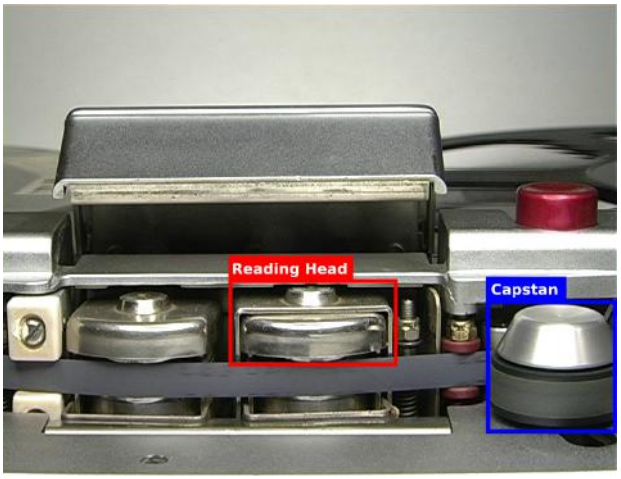

FIGURE 4. Tape machine Reading Head and Capstan.

Once ROIs are identified, the Video Analyzer proceeds to detect irregularities within the digitized magnetic tape images. This is accomplished by examining the absolute value of the differences between consecutive frames in grayscale. Specifically, the function defined by Equation 1 generates a new grayscale image of the tape based on the difference between consecutive input frames.

$$\mathbf{D}(i,j) = |\mathbf{C}(i,j) - \mathbf{P}(i,j)| \tag{1}$$

where $i = 1,\ldots,n$ and n is the number of rows in the matrix, $j = 1,\ldots,m$ and m is the number of columns in the matrix, matrix $\mathbf{D}$ is the difference frame, $\mathbf{C}$ is the current frame matrix and $\mathbf{P}$ is the previous frame matrix in grayscale.

TABLE I. Standard deviation S based on tape's speed

Speed (ips)	Empirical standard deviation
30	2.25
15	2.5
7.5	2.6
3.75	2.75

Given the tape's potential color variations, the standard deviation of the difference image's color is considered to enhance algorithm stability in the presence of color and brightness fluctuations. Instead of merely using the mean, both mean and standard deviation of the pixel color in grayscale are computed using Equations 2 and 3, respectively. These metrics are then compared with the estimated standard deviation (S) based on the tape's speed, as summarized in Table I, where the estimated standard deviation has been calculated through empirical tests. The tests involved examining 30 video frames without irregularities for each considered tape speed. For each frame, the mean color value was calculated, and subsequently, the mean and standard deviation were computed.

The following two equations are defined to calculate the mean and standard deviation of the pixel's color (in grayscale, so a single 8-bit channel) of an image $\mathbf{D}$ of dimension $m \times n$:

$$\mu = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\mathbf{D}(i,j) \tag{2}$$

$$\sigma = \sqrt{\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\big(\mathbf{D}(i,j) - \mu\big)^2} \tag{3}$$

where $\mathbf{D}(i,j), m, n$ have the same meaning as defined in Equation 1.

When $\sigma < S$, it indicates that the colors in the image show minimal visible variation, implying that the difference image does not contain any anomalies. Otherwise, for irregular difference images, the Otsu thresholding method [33] is applied to define a threshold (T) for converting the image to a binary format using Equation 4 and Equation 5.

$$T = arg\,max_t\{\sigma_B^2(t)\cdot w_B(t) + \sigma_F^2(t)\cdot w_F(t)\} \tag{4}$$

where $\sigma_B^2(t)$ is the weighted variance of the class above the threshold, $w_B(t)$ is the probability of the class above the threshold, $\sigma_F^2(t)$ is the weighted variance of the class below the threshold and $w_F(t)$ is the probability of the class below the threshold. Then thresholding is applied to the image using Equation 5.

$$\mathbf{B}(i,j) = \begin{cases} 1 & \text{if } \mathbf{D}(i,j) \geq T \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where $\mathbf{B}$ is the binary image obtained after thresholding.

$$\mathbf{O}(i,j) = \mathbf{B}(i,j) \circ SE = (\mathbf{B}(i,j) \ominus SE) \oplus SE \tag{6}$$

where SE is the structuring element defined as a 3×3 squared matrix.

$$I = \begin{cases} 1 & \text{if } \sum_{i=0}^{m}\sum_{j=0}^{n}\mathbf{O}(i,j) > \frac{m\times n\times 5}{100} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Subsequently, a denoising function is applied to highlight irregularity shapes through opening operations (erosion and dilation) with a 3×3 rectangular kernel, as described in Equation 6. After this process the matrix $\mathbf{O}(i,j)$ should contain a clearer shape of the irregularity. The count of white pixels in the resulting image is computed, and if it exceeds 5% of the image's area, a significant difference between consecutive frames is inferred. Equation 7 summarizes the decision process: if the area of difference in the image exceeds a fixed threshold, it is considered an irregularity.

At the end of the detection process, the frames in which an irregularity was found are stored as Irregularity Images along with their timestamp and a unique id in a JSON file called *IrregularityFile*.

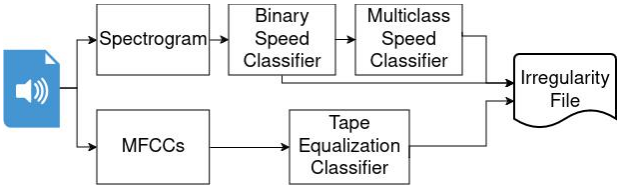

FIGURE 5. Components of the Audio Analyzer.

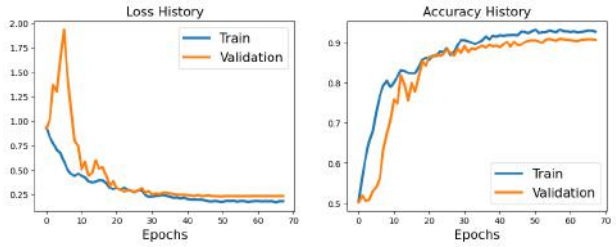

FIGURE 6. DNN Equalization classifier training curve.

2) Audio Analyzer

The Audio Analyzer AIM is responsible for carrying out a spectral analysis of the Preservation Audio File, identifying playback equalization requirements, detecting tape speed errors and computing the cross-correlation between the high-quality preservation audio track and the lower-resolution audio track in the associated video file. This process is vital for achieving synchronization between the Preservation Audio File and the Preservation Audio-Visual File, ensuring the alignment and coherence of the audio components during the preservation process. A previous approach, as described in [24], relied on a single classifier to identify all audio irregularities. A novel approach, presented for the first time in this paper, divides the signal classification into three distinct phases. In the first two phases, machine learning methods are employed to respectively recognize equalization curves and reading/writing tape speeds. The third phase focuses on computing the signal cross-correlation between the audio and video components. This approach enhances efficiency and accuracy in handling the diverse aspects of signal irregularities detected during the preservation process.

Expanding on prior research detailed in [34] and [35], our current spectral analysis utilizes the first 13 Mel-Frequency Cepstral Coefficients (MFCCs) to represent audio for equalization classification (see lower part of Fig. 5). This representation proves to be a suitable approximation of the signal for the task of equalization classification. Given the significant variability of the content recorded on the tape, we decided to adhere to prior research by concentrating the analysis solely on segments of silence on the tape, i.e., portion of the signal with intensity below -50 dBFS. Silence, as defined in [35], where empirical findings indicate that audio signals with intensities ranging from -50 to -63 dBFS represent silence between spoken words, from -63 to -69 dBFS represent noise resulting from the recording head without input, and below -69 dBFS represent noise from sections of pristine tape, tends to produce more consistent results when employed for classification, as opposed to analyzing the entire signal present on the tape. The dataset under analysis comprises 9328 audio segments, each lasting 500 ms and featuring intensities below -50 dBFS. From these segments, the first 13 Mel-Frequency Cepstral Coefficients (MFCCs) are extracted. These segments are part of a collection of 25 audio tape recordings, designed to encompass every possible configuration of IEC1 and IEC2 equalization curves at different tape speeds. These tapes were digitized at a sampling rate of 96 kHz and a bit depth of 24 bits. The classification process involved the dataset preparation, the model selection and an assessment over the test set. The performance of all the models has been validated and tested using a K-Fold cross-validation with $k = 5$ and a $80, 20$ train-validation dataset split. Numerous experiments were conducted performing grid search cross-validation over K-Nearest Neighbour (KNN), Random Forest Classifiers (RFC), Support Vector Machines (SVM) and Gradient Boosting (XGB) to tune the algorithms hyperparameters and a Deep Neural Networks (DNN) whose structure is described in the following paragraph. All models were trained and validated using the same dataset to select the most effective model. As illustrated in Table II, the best results over the test set were achieved using DNNs.

TABLE II. Models scores over test set in EQ recognition

Model	Accuracy	Precision	Recall
KNN	0.80	0.79	0.76
RFC	0.84	0.86	0.82
SVM	0.88	0.88	0.87
XGB	0.87	0.87	0.87
DNN	0.90	0.90	0.90

The constructed neural network architecture comprises eleven fully connected layers. The input layer matches the dimensions of the input data, followed by two layers each containing 128 neurons, two layers with 64 neurons, two layers with 32 neurons, and two layers with 8 neurons. A dropout layer is included to mitigate overfitting, followed by an output layer comprising three neurons. The network encompasses approximately 37,000 parameters in total. Activation functions employed within layers containing 128, 64, 32, and 8 neurons consist of Leaky ReLU, complemented by batch normalization techniques to counteract potential issues arising from vanishing gradients.

Loss and accuracy metrics derived from the validation and training datasets are presented in Fig. 6. Notably, the absence of conspicuous overfitting is evident, as indicated by the consistent behavior of the validation curve across epochs. However, it is discernible from the figure that the model performance tends to plateau after a certain number of epochs. Consequently, the number of training epochs has been constrained to 50 epochs to optimize computational efficiency while maintaining satisfactory performance levels.

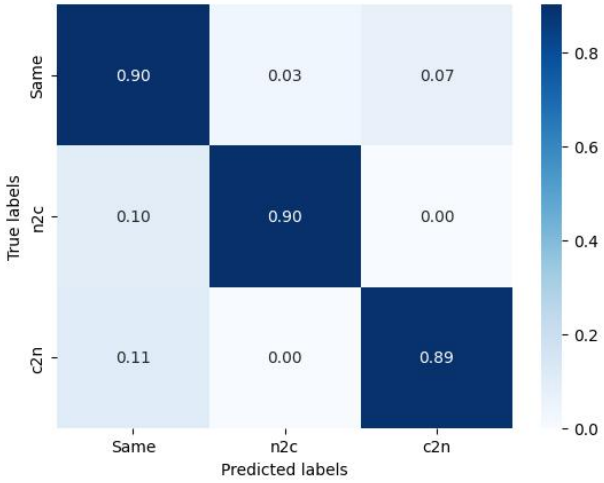

FIGURE 7. Equalization classifier normalized confusion matrix evaluating performance over the test set, "same" label means that writing and reading equalization were the same, n2c means that the tape have been read in IEC1 and written in IEC2, while c2n is the opposite.

A final evaluation of the model can be made by observing the confusion matrix in Fig. 7: although the major diagonal of the matrix can be clearly seen, the results highlight that, generally, when the classifier returns an incorrect result, the digitized tapes with different pre and post emphasis equalization applied during recording and playback are recognized as correctly processed tapes. It must be remembered, however, that in this phase the individual 500 ms audio segments are treated separately, while the subsequent modules of the ARP AIW deal with aggregating the information, therefore the problem of the wrong classification result should be mitigated later in the ARP AIW. In fact, a single classification mistake in the middle of a sequence of many correctly classified segments does not influence the outcome of the preservation process.

The audio playback speed detection algorithm is based upon the analysis of spectrogram images of various audio files (see upper part of Fig. 5). These images provide a visual representation of the variations of the audio spectrum over time. In this case the dataset consists of 300 audio files.

The purpose of compiling this dataset was to encompass sounds with a wide variety of spectral characteristics.

The audio files are categorized into different groups: those with correct playback speed and those with speed variations. Since tape speed varies by factors of 2, the speed changes have been labeled with relative changes rather than absolute values. In other words, an audio file with a change in speed from 3.75 ips to 7.5 ips is in the same category (double) as an audio file with a change in speed from 7.5 to 15 ips. The same applies for halving the speed. Since the most common speeds used in professional audio recordings are 3.75, 7.5, 15 ips, the identified categories are: double when the tape is read at double the writing speed, half when the tape is read at half the writing speed writing, quarter if the tape is read at a quarter of the writing speed and quadruple as the reciprocal of the previous case.

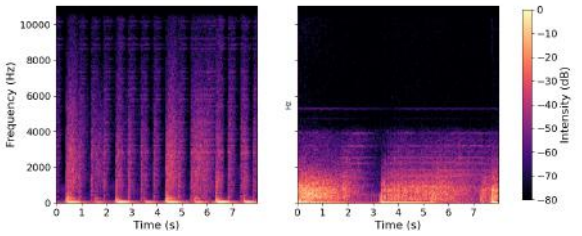

FIGURE 8. Audio spectrogram images showing drum tape recordings played back at the correct speed, 15 ips (left), and at a quarter of the correct speed, 3.75 ips (right).

The audio spectrogram images, similar to those shown in Fig. 8, were extracted separately for each audio channel. Initially, the spectrogram of the entire audio file was generated and then split into chunks of 1 s duration in grayscale at 8 bits, with dimensions of 256×128 pixels. Additionally, durations of 250 ms (64×128 pixels) and 500 ms (128×128 pixels) were tested, but they yielded poorer results as displayed in Table III.

TABLE III. Model scores for audio segments with varying durations

Input Size	F1-score	Precision	Recall
Binary model			
64×128	0.77	0.80	0.79
128×128	0.94	0.94	0.94
256×128	0.96	0.96	0.96
Multiclass model			
64×128	0.40	0.71	0.45
128×128	0.97	0.97	0.97
256×128	0.98	0.98	0.98

To minimize errors, the speed classification task has been divided into two stages. The first is a binary classifier that determines if the tape is correctly played or not. The second classifier is activated only in case of anomaly detection and is used to determine the error class (double, half, quarter, quadruple). Both stages utilize Convolutional Neural Network (CNN) models. The two models share the same structure, apart from the last layer responsible for outputting the predicted label. Each model comprises three convolutional layers with a kernel size of 7×7 and a ReLU activation function. The layers differ in the number of neurons, progressively increasing from 8, to 16, and finally to 32. Each convolutional layer is followed by a max pooling layer with kernel 5×5. After these layers, global average pooling is performed, and its output connects to a dense layer of size 32, always with a ReLU activation function. The final layer consists of 2 neurons for the binary classifier, activated with the sigmoid function, and 4 neurons for the multi-class classifier, using the softmax activation function.

To evaluate the performance of the playback speed detection models, extensive testing and validation were conducted using a diverse set of audio recordings with known speed

variations. Both stages achieved impressive accuracy scores, with precision, recall, and F1-score values exceeding 98% on the validation set, while the test set gave a noticeable performance drop around to 85%. As illustrated in Fig. 5, the output of the Audio Analyzer consists of the Irregularity File, which includes detected irregularities metadata collected in a single JSON file.

D. TAPE IRREGULARITY CLASSIFIER

The Tape Irregularity Classifier is designed to verify and merge, if necessary, the irregularities received as input from the Audio and Video Analyzer AIMs through the irregularity files (see Fig. 11). It utilizes a CNN model tailored specifically for analyzing irregularities extracted from video data and consolidates the classifications related to the individual audio chunks.

In the current implementation the Classifier CNN model is trained to recognize three classes of Irregularities on the tape: Splices, Brands, and Shadows. While splices constitute the primary focus in video analysis, brands and shadows serve to accommodate detections made by the video analyzer that are not strictly content-related. Brands marks on the tape, though prevalent, lack relevance to audio and metadata content. While they recur consistently throughout the tape, the brand information is stored only once, with subsequent brand images segregated into a distinct folder and excluded from the irregularity file. Shadows, conversely, may arise due to specific lighting conditions or irregularities on the tape surface. The latter scenario is of paramount importance in preservation endeavors, necessitating the retention of shadows as irregularities in the Irregularity File to prevent information loss.

Initially, our approach involved leveraging transfer learning by fine-tuning a pre-trained model based on EfficientNet B0 [36], which has shown effectiveness across various computer vision tasks. The classifier architecture consists of an input layer accepting 224×224 pixel color images, followed by convolutional layers with frozen weights responsible for extracting relevant features from the input data. A Global Average Pooling layer combined with a dense layer forms the output, with the number of neurons in the dense layer corresponding to the number of irregularity classes to be recognized (in this instance, $n = 3$). Notably, the EfficientNet model accepts images with 8-bit values instead of pixel values scaled between 0 and 1.

TABLE IV. Models scores over test set in Irregularity images recognition

Model	Accuracy	Precision	Recall
EfficientNet B0	0.91	0.90	0.91
Custom Speed Classifier	0.96	0.96	0.96

In addition to the initial approach that leveraged transfer learning with the EfficientNet architecture, we conducted another experiment using an adapted version of the CNN architecture as detailed in the preceding section, the Speed Change Classifier network. The former showed excellent results in both training and validation, achieving approximately 99% accuracy. Its performance, however, dropped by approximately 10% on the test set, suggesting potential overfitting. In contrast, the latter model, with slightly lower accuracy during training and validation (around 97%), demonstrated better generalization on the test set with an accuracy of nearly 96%. The summarized results can be found in Table IV.

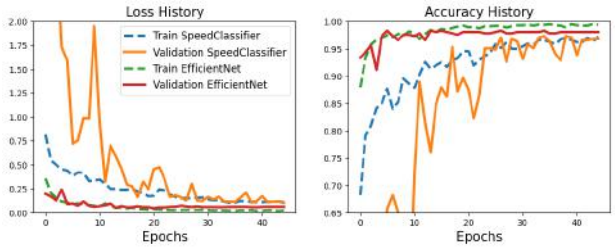

FIGURE 9. Tape Irregularity Classifier training curve.

This suggests that the Speed Change Classifier network architecture offers a more stable performance, highlighting its potential for broader applicability in analyzing tape irregularities. In fact, as can be seen from Fig. 9, EfficientNet's accuracy on validation set tends to overfit from the early epochs, whereas the other's model accuracy score exhibits slightly more variability across epochs but demonstrates a consistent improvement with an increase in the number of training epochs.

The training dataset comprises Irregularity Images corresponding to each class. The training set is partitioned into 80% for training and 20% for validation. Subsequently, the model undergoes testing on Irregularity Images detected from recently digitized magnetic audio tapes, ensuring evaluation on a distinct set of images not encountered during training. This approach facilitates assessing the model's ability to generalize to unseen data and accurately identify irregularities across disparate sources. The dataset, overall, exhibits slight class imbalance, with splices comprising approximately 800 images, while brands and shadows each contain around 600 images. Following 20 epochs of training, the model demonstrates notable efficacy, achieving a 97% accuracy over the validation dataset, indicative of its adeptness in discerning patterns associated with splices, brands, and shadows within the provided dataset.

Looking at the confusion matrix of the model (see Fig. 10), no particular class imbalances emerge in the results, Brands and Splices are the classes that reap the greatest successes, while Shadows are occasionally confused with the other classes.

Upon completion of the irregularities selection and aggregation process, the Tape Irregularities Classifier AIM shares the chosen Irregularity Images (along with their metadata) with the Packager. Simultaneously, the aggregated audio irregularities are transmitted to the Tape Audio Restoration AIM to allow the generation of a restored Access Copy. The specific data flow is depicted in Fig. 11.

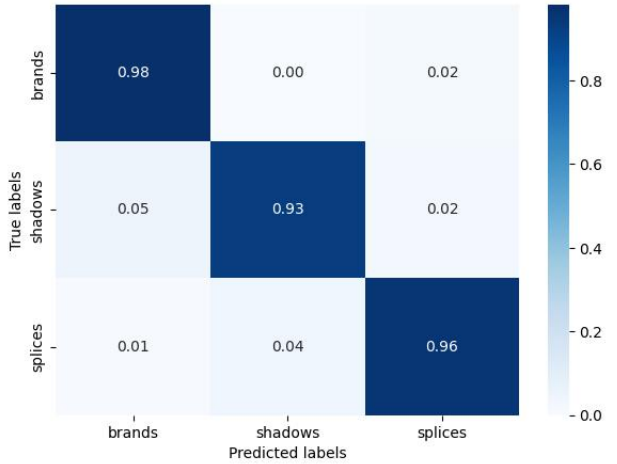

FIGURE 10. Tape Irregularity Classifier normalized confusion matrix evaluation performance over test set.

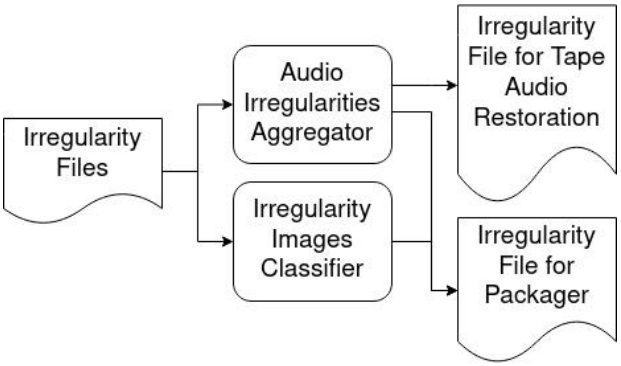

FIGURE 11. Tape Irregularity Classifier data flow.

E. TAPE AUDIO RESTORATION

The Tape Audio Restoration AIM tackles audio irregularities by rectifying time-reversed segments, conducting sampling rate conversion for accurate playback speed, and applying equalization correction curves in areas identified by the Irregularity File. Inputs to this AIM include the Irregularity File and the Preservation Audio File. The resulting Restored Audio Files guarantee the appropriate playback of the original audio content.

The current implementation for the equalization curve correction is based on [37]. The equalization correction workflow operates in two steps: first it applies the inverse of the curve used during the digitization process and then it applies the correct equalization curve. The correction of speed happens through resampling where the new sample rate (sr) is computed as follow:

$$sr_{new} = sr_{old} \frac{speed_{writing}}{speed_{reading}} \tag{8}$$

Finally, the correction of reversed audio is implemented by simply reversing in time the order of the audio samples based on the starting and ending points of the detected irregularity.

F. PACKAGER

Once all the metadata, magnetic tape images, and restored audio segments are obtained, it is crucial to provide easily accessible and searchable files. The *Packager* AIM is responsible for this task, receiving all the materials generated by the other AIMs and organizing them into folders. One folder is designated for storing a philological copy of the tape (audio and video, synchronized, in high resolution without any corrections) along with the metadata of the identified irregularities (Preservation Master Files). Another folder is created to provide access to a (potentially) restored audio file, sometimes in a compressed format, making it easy to download and play on various devices (Access Copy Files).

V. CONCLUSIONS

This paper introduces the innovative technology integrated into the ARP standard, showcasing exceptional results in the digital preservation and restoration of magnetic tapes. The test outcomes for the Tape Irregularity Classifier AIM reveal that custom neural network architectures remain relevant and can effectively compete with established CNNs. This indicates that task-specific models can outperform more general ones in certain scenarios. Following rigorous training and testing, the Tape Irregularity Classifier achieves an impressive 96% accuracy on the test dataset, demonstrating its ability to generalize to new data and accurately detect irregularities across various sources. Moreover, the classifier's robustness is evident in its equitable handling of class imbalances within the dataset, ensuring unbiased recognition of different irregularity types.

Similarly, the Audio and Video Analyzer AIMs demonstrate exceptional performance in identifying and characterizing tape irregularities. The Video Analyzer, employing sophisticated techniques such as ROI identification and difference frame analysis, accurately detects anomalies while mitigating false positives arising from environmental variations. By focusing the video on stationary elements in the tape playback system and employing advanced image processing algorithms, the Video Analyzer ensures reliable identification of irregularities, essential for preserving valuable annotations stored on magnetic tapes. The Audio Analyzer AIM, employing meticulous feature extraction and model selection, attains exceptional performance in identifying equalization curves and playback speeds. This highlights the model's ability to classify audio irregularities with high accuracy, thereby aiding in the comprehensive preservation of audio content archived on magnetic tapes.

In conclusion, the advanced technology integrated into the ARP standard offers effective solutions for detecting and characterizing irregularities, contributing to the preservation of valuable audio records. Adoption of the CAE-ARP standard empowers archives to efficiently identify and rectify errors in various audio files, improving the quality and accuracy of preservation and access copies, streamlining archiving processes, and ensuring interoperability through standardized digital file formats. The ARP standard marks a

significant advancement in the preservation of audio cultural heritage, ensuring its enduring accessibility and usability, and paving the way for future developments in this critical field.

REFERENCES

[1] C. Woideck, ed., Charlie Parker: His Music & Life: His Music and Life. University of Michigan Press, 1998.
[2] S. Canazza and G. De Poli, "Four decades of music research, creation, and education at Padua's Centro di Sonologia Computazionale," Computer Music Journal, vol. 43, no. 4, pp. 58–80, 2020.
[3] S. Canazza, G. De Poli, and A. Vidolin, "Gesture, music and computer: The centro di sonologia computazionale at padova university, a 50-year history," Sensors, vol. 22, no. 9, 2022.
[4] 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, vol. 8, 2018.
[5] N. Pretto, A. Russo, F. Bressan, V. Burini, A. Rodà, and S. Canazza, "Active preservation of analogue audio documents: a summary of the last seven years of digitization at CSC," in Proceedings of the 17th sound and music computing conference 2020, SMC'20, Torino, Italy, 2020.
[6] K. Bradley, ed., IASA-TC 04, IASA Technical Committee, Guidelines on the Production and Preservation of Digital Audio Objects. www.iasa-web.org/tc04/audio-preservation, 2006.
[7] M. Miller and S. Gherdevic, Guidelines for Audiovisual and Multimedia Collection Management in Libraries. International Federation of Library Associations and Institutions (IFLA), IFLA Audiovisual and Multimedia Section Standing Committee, 2017.
[8] F. Bressan and S. Canazza, "A systemic approach to the preservation of audio documents: Methodology and software tools," Journal of Electrical and Computer Engineering, vol. 2013, p. 21 pages, 2013.
[9] M. Mansoori and R. Murali, "A systematic survey on music composition using artificial intelligence," in 2022 International Conference for Advancement in Technology (ICONAT), pp. 1–8, 2022.
[10] H. Cai, T. Pu, Y. Luo, and X. Zhou, "Music genre prediction based on machine learning," in 2021 IEEE International Conference on Artificial Intelligence and Industrial Design (AIID), pp. 198–201, 2021.
[11] M. Catak, S. AlRasheedi, N. AlAli, G. AlQallaf, M. AlMeri, and B. Ali, "Artificial intelligence composer," in 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), pp. 608–613, 2021.
[12] J.-W. Hwang, R.-H. Park, and H.-M. Park, "Efficient audio-visual speech enhancement using deep u-net with early fusion of audio and video information and rnn attention blocks," IEEE Access, vol. 9, pp. 137584–137598, 2021.
[13] T. Saeki, S. Takamichi, T. Nakamura, N. Tanji, and H. Saruwatari, "Selfremaster: Self-supervised speech restoration for historical audio resources," IEEE Access, vol. 11, pp. 144831–144843, 2023.
[14] A. Ragano, E. Benetos, and A. Hines, "Automatic quality assessment of digitized and restored sound archives," JAES Volume 70 Issue, 2022.
[15] F. Miotello, M. Pezzoli, L. Comanducci, F. Antonacci, and A. Sarti, "Deep prior-based audio inpainting using multi-resolution harmonic convolutional neural networks," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, pp. 113–123, 2024.
[16] G. Colavizza, T. Blanke, C. Jeurgens, and J. Noordegraaf, "Archives and ai: An overview of current debates and future perspectives," J. Comput. Cult. Herit., vol. 15, dec 2021.
[17] R. L. Hess, "Tape degradation factors and challenges in predicting tape life," ARSC Journal, vol. 39, p. 240, 2008.
[18] F. Bressan and S. Canazza, ""Honey, i burnt the tapes!" a study on thermal treatment for the recovery of magnetic tapes affected by soft binder syndrome-sticky shed syndrome," IASA Journal, vol. 44, pp. 53–64, January 2015.
[19] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, and M. Dietz, "ISO/IEC MPEG-2 advanced audio coding," J. Audio Eng. Soc, vol. 45, no. 10, pp. 789–814, 1997.
[20] M. Bosi and R. Goldberg, Introduction to Digital Audio Coding and Standards. Kluwer Academic Publishers - Springer, 2002.
[21] Magnetic tape sound recording and reproducing systems. IEC - International Electrotechnical Commission, 1981.
[22] Magnetic Tape Recording and Reproducing Standard, Reel-to-Reel. NAB - National Association of Broadcasters, 1965.
[23] M. Bosi, N. Pretto, M. Guarise, and S. Canazza, "Sound and music computing using AI: Designing a standard," in Proceedings of the 18th Sound and Music Computing Conference 2021, SMC'21, Virtual Conference, 2021.
[24] M. Bosi, S. Canazza, A. Russo, N. Pretto, and L. Chiariglione, "An MPAI/IEEE International Standard for Audio: Overview of CAE Audio Recording Preservation (ARP) Technology," in Audio Engineering Society Conference: 2023 AES International Conference on Audio Archiving, Preservation & Restoration, Jun 2023.
[25] A. Basso, P. Ribeca, M. Bosi, N. Pretto, G. Chollet, M. Guarise, M. Choi, F. Yassa, and R. Iacoviello, "Ai-based media coding standards," SMPTE Motion Imaging Journal, vol. 131, no. 4, pp. 10–20, 2022.
[26] A. D. Birrell and B. J. Nelson, "Implementing remote procedure calls," ACM Trans. Comput. Syst., vol. 2, p. 39–59, feb 1984.
[27] N. Pretto, E. Micheloni, A. Chmiel, N. D. Pozza, D. Marinello, E. Schubert, and S. Canazza, "Multimedia archives: New digital filters to correct equalization errors on digitized audio tapes," Advances in Multimedia (Hindawi), vol. 2021, p. 5410218, 2021.
[28] C. Fantozzi, F. Bressan, N. Pretto, and S. Canazza, "Tape music archives: from preservation to access," International Journal of Digital Libraries, vol. 18, no. 233, 2017.
[29] A. Russo, M. Spanio, and S. Canazza, "Enhancing preservation and restoration of open reel audio tapes through computer vision," in Image Analysis and Processing - ICIAP 2023 Workshops (G. L. Foresti, A. Fusiello, and E. Hancock, eds.), (Cham), pp. 297–308, Springer Nature Switzerland, 2024.
[30] D. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," Pattern Recognition, vol. 13, no. 2, pp. 111–122, 1981.
[31] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," Computer Vision and Image Understanding, vol. 110, no. 3, pp. 346–359, 2008. Similarity Matching in Computer Vision and Multimedia.
[32] G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000.
[33] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62–66, 1979.
[34] 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 AI*CH 2017. The 11th workshop on Artificial Intelligence for Cultural Heritage. In CEUR WORKSHOP PROCEEDINGS, vol. 2034, pp. 17–24, 2017.
[35] N. Pretto, C. Fantozzi, E. Micheloni, V. Burini, and S. Canazza, "Computing methodologies supporting the preservation of electroacoustic music from analog magnetic tape," Computer Music Journal, vol. 42, no. 4, pp. 59–74, 2019.
[36] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," CoRR, vol. abs/1905.11946, 2019.
[37] N. Pretto, N. d. Pozza, A. Padoan, A. Chmiel, K. J. Werner, A. Micalizzi, E. Schubert, A. Roda, S. Milani, and S. Canazza, "A workflow and digital filters for correcting speed and equalization errors on digitized audio open-reel magnetic tapes," Journal of the Audio Engineering Society, vol. 70, pp. 495–509, june 2022.

MARINA BOSI , a pioneer in digital audio coding, is a founding director of MPAI and chairs the IEEE SA CAE WG. A fellow and past president of the Audio Engineering Society, Dr. Bosi was chief technology officer of MPEG LA, LLC, Denver, CO, vice president-technology at DTS, Inc., Los Angeles, CA, and was a member of the research team that created Dolby Digital at Dolby Laboratories, San Francisco, CA, where she also led the MPEG-2 AAC development for which she received the ISO/IEC 1997 Editor Award. Dr. Bosi holds a degree in Physics from the University of Florence, completing her dissertation at IRCAM in Paris, and a degree from the Conservatory of Florence, having later served as a faculty member at the Conservatory of Venice. She's a graduate of Stanford Business School's Executive Program and has held fiduciary positions on several boards. A sought-after keynote speaker, Dr. Bosi holds multiple patents and authored significant contributions to academic literature, including the textbook, Introduction to Digital Audio Coding and Standards (Kluwer/Springer,2002). In recognition of her achievements, Dr. Bosi has received numerous awards, including the AES Silver Medal.

SERGIO CANAZZA is professor of: "Fundamental of Computer Science" and "Computer Engineering for Music and Multimedia", Department of Information Engineering, University of Padova. He is scientific director of the Centro di Sonologia Computazionale, and his main research interests involve 1) Expressive information processing, 2) Auditory displays, and 3) musical cultural heritage preservation and exploitation. He is author or co-author of more than 250 articles on International Journals and Refereed International Conferences. He has been i) general chairman and member of Technical Committees at several conferences and ii) Project Manager in European projects. He is CEO of Audio Innova srl, an university spin-off enterprise (founder member of MPAI), and he is owner of patents on safety and health at work. He won Le Palm D'Or at the World Artificial Intelligence Cannes Festival (France) in 2023 and in 2024, one of the most important AI event at the world level.

NICCOLÒ PRETTO completed his Bachelor's and Master's degrees in Computer Science Engineering and his PhD in Information Engineering at the Department of Information Engineering, University of Padova, Italy. He is an Assistant Professor at the Free University of Bozen-Bolzano, Italy, and he is part of the Media Interaction Lab. His research is primarily focused on sound and music computing, and preservation and access to historical audio documents and cultural heritage in general. More specifically, his work consists of research and development of innovative methodologies, and applications to preserve, analyze, and experience musical cultural heritage, adopting several technologies and methods extending to web and mobile interfaces, embedded systems, and machine learning techniques. He is also a founder member of the company MPAI Store Limited.

ALESSANDRO RUSSO got a Bachelor's degree in Technologies for Cultural Heritage at the University of Turin in 2012 and a Master's degree in Materials Science for Cultural Heritage in 2015. His main research activities concern the preservation of audio, film, and video archives. Since 2016, he has carried out research activities collaborating with the Centro di Sonologia Computazionale (CSC) of the Department of Information Engineering (DEI) of the University of Padua, mainly focusing on audio restoration, and with La Camera Ottica Lab, Film and Video Restoration Lab of the University of Udine, working on several projects of digitization and restoration of audio-visual funds belonging to Italian and International foundations and archives. Currently, he is a Ph.D. student in *Brain, Mind, and Computer Science* at the University of Padua, with a project concerning the preservation, re-activation, and documentation of audio-visuals and Multimedia Installations.

MATTEO SPANIO earned his Bachelor's degree in Computer Science with a specialization in Data Science from Ca' Foscari University of Venice. He is currently a Ph.D. student in the *Brain, Mind, and Computer Science* program at the University of Padova. His research interests concern the intersection of artificial intelligence and music, focusing on generative AI for music based on cross-modal interaction and the preservation of audio documents. In addition to his academic pursuits, he collaborates as a software engineer with the companies Soundfood and Audio Innova. Furthermore, he holds Bachelor's and Master's degrees in Performing Arts from the Conservatorio di Musica "C. Pollini" in Padua and he frequently performs as first clarinet in many professional orchestras and has graced prestigious stages and theaters across Italy, Hungary, Austria, and Germany.