

## Research Article

# Multimedia Archives: New Digital Filters to Correct Equalization Errors on Digitized Audio Tapes

Niccolò Pretto <sup>1</sup>, Edoardo Micheloni,<sup>1</sup> Anthony Chmiel <sup>2</sup>, Nadir Dalla Pozza,<sup>1</sup>  
Dario Marinello,<sup>1</sup> Emery Schubert,<sup>3</sup> and Sergio Canazza<sup>2</sup>

<sup>1</sup>Department of Information Engineering, University of Padova, Padova, Italy

<sup>2</sup>The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia

<sup>3</sup>Empirical Musicology Laboratory, University of New South Wales, Sydney, Australia

Correspondence should be addressed to Niccolò Pretto; [niccolo.pretto@dei.unipd.it](mailto:niccolo.pretto@dei.unipd.it)

Received 8 November 2019; Revised 15 February 2021; Accepted 17 March 2021; Published 30 March 2021

Academic Editor: Patrick Seeling

Copyright © 2021 Niccolò Pretto et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Multimedia archives face the problem of obsolescing and degrading analogue media (e.g., speech and music recordings and video art). In response, researchers in the field have recently begun studying *ad hoc* tools for the preservation and access of historical analogue documents. This paper investigates the active preservation process of audio tape recordings, specifically focusing on possible means for compensating equalization errors introduced in the digitization process. If the accuracy of corrective equalization filters is validated, an archivist or musicologist would be able to experience the audio as a historically authentic document such that their listening experience would not require the recovery of the original analogue audio document or the redigitization of the audio. Thus, we conducted a MUSHRA-inspired perception test ( $n = 14$ ) containing 6 excerpts of electronic music (3 stimuli recorded NAB and 3 recorded CCIR). Participants listened to 6 different equalization filters for each stimulus and rated them in terms of similarity. Filters included a correctly digitized “Reference,” an intentionally incorrect “Foil” filter, and a subsequent digital correction of the Foil filter that was produced with a MATLAB script. When stimuli were collapsed according to their filter type (NAB or CCIR), no significant differences were observed between the Reference and MATLAB correction filters. As such, the digital correction appears to be a promising method for compensation of equalization errors although future study is recommended, specifically containing an increased sample size and additional correction filters for comparison.

## 1. Introduction

The transition required for the information age brings with it the need to transfer preexisting (analogue) multimedia materials into a digital form in order to withstand the wear and tear of time and the progression of technology, such as search and recovery functions through increasingly powerful digital tools. Archiving has become an increasingly important goal both in terms of historical documentation and also for ease of location and availability. The implications of these needs are particularly complex when it comes to historical music recordings. In this context, research on the preservation and restoration of sound documents has been developed in the information engineering area and, in particular, in the multimedia field, augmenting the

innovations introduced for storage and retrieval technologies [1]. These developments have additional implications for the definition of digitization protocols to help ensure maintenance and longevity.

This paper presents the problem of equalization in the active preservation process of audio documents. If the goal of the active preservation and re-recording process is to pursue historical faithfulness, the audio signal must be precisely filtered to take into consideration recording equalization that is part of the original source audio document [2]. Choosing the correct equalization curve is essential to avoid the proliferation of additional, incorrect versions of the audio documents (referred to in philology as a “false witnesses” [3]). The choice is usually made on the basis of both historical information (which is rarely complete

and exhaustive) and the experience of the technicians [4], introducing a certain margin of interpretation. We therefore present tools to compensate for errors (in choosing the equalization curve) introduced by the re-recording technicians. In this way, if an archivist or musicologist notices that a preservation master has been produced using the wrong equalization curve, it can be changed without having to recover the original analogue audio document (which may have deteriorated in the meantime).

In Section 2, we present an overview of analogue equalization, illustrating the problems concerned with the user choice. Next, we focus on a case study of the analogue audio tape and explain two equalization standards from a mathematical point of view. In Section 3, these equalizations will be transformed into the digital domain, and in Section 4, we report an experiment assessing the perception of these equalization methods. Based on the results, we propose that a digital correction filter provides a reliable means to compensate for errors made in the digitization process.

## 2. Analogue Equalizations

*2.1. The “Equalization Problem”.* The term “equalization” can be used to indicate any procedure that involves altering or adjusting of the overall frequency spectrum characteristics of the audio signal. 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 analogue audio recordings [5]. In analogue audio recordings, a preemphasis curve is applied to the signal which is contained in the analogue carrier, and an inverse post-emphasis curve is applied during the reproducing phase. Thus, the resulting output signal maintains nearly the flat frequency response of the original input [6], but at the same time, it is characterized by an extension of the dynamic range [7] and an improvement of the SNR [8]. This technique is adopted from several analogue audio technologies due to the limited dynamic range of audio systems [7].

Historically, the adoption of these techniques was not uniform, and several different standards were applied by record manufacturers. To faithfully reproduce recordings, it is necessary to tackle what is referred to as the “equalization problem [9].” This problem specifically arises when analyzing magnetic tape technology. Several standards exist [4], and during playback and digitization, this must be considered to help obtain an “authentic” listening experience, that is, postemphasis filtering (equalization) that corresponds to that of the machines upon which the playback was originally intended. The differences between the equalization curves are subtle, and during the digitization process, it may be difficult to determine the “correct” one, and without reliable documentation or test tones, operators involved in the digitization process are forced to choose the equalization aurally [4, 9], which may lead to errors. Therefore, there is a possibility that an “incorrect” equalization will be selected in the process of digitizing audio tapes. These issues could be resolved through innovative automatic analysis tools, as recently presented in [10], or through an accurate historical investigation of the recording studio, aiming to individuate

the original equipment and the relative setup used at the time [11].

The musicological study of sound recording is often performed directly on the digital copy. If, at this stage, the musicologist has doubts about the type of equalization used during the analogue-digital transfer, it is beneficial to provide her/him with corrective tools, which enable comparisons between the existing, possibly inauthentic versions and corrected versions. It is not feasible to redigitize audio tapes with the correct equalization on a large scale due to excessive economic cost of the operation. Furthermore, a number of these heritage items may now be unreadable due to physical degradation [2], making the matter of corrective equalization an urgent one. The solution proposed in this paper is to create a set of precise digital filters to subtract the “incorrect” equalization curve applied in the digitization process and to add a corrective measure. It is important to specify that these filters must only be used to alter access copies or with access tools such as those presented in [12, 13] that filter the signal without performing irreversible changes to the file. That is, they must not alter the preservation copy for any reason.

*2.2. Case Study.* 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 the United States. 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 the United States, the most prevalent standard was IEC2, also referred to as NAB from the American National Association of Broadcasters. We henceforth refer to these as CCIR and NAB, being the two standards that this paper focuses on. The equalization standards are strictly connected to another parameter that must be correctly configured before the equalization setting: the playback speed. There are 6 standard speeds, but the most common are 15 ips (38.1 cm/s) and 7.5 ips (19.05 cm/s) [4]. The latter speed will be used in our work. As can be seen in [14], digitization problems derived by different speed (and therefore equalization) standards in the same open-reel tape are quite widespread. Nevertheless, in this preliminary study, the authors decided to not involve a second variable. Further study will be necessary for correcting both speed and equalization errors.

The first step of this work consists of the analysis of the pre- and postemphasis curves for any standard. A post-emphasis curve could be expressed as a combination of two curves described with the following formula:

$$N(\text{dB}) = 10 \log\left(1 + \frac{1}{4\pi^2 f^2 t_1^2}\right) - 10 \log(1 + 4\pi^2 f^2 t_2^2), \quad (1)$$

where  $f$  is the frequency in Hz and  $t_1$  and  $t_2$  are the time constants in microseconds [15]. An alternative mathematical representation of the formula is

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

where  $\omega = 2\pi f$ , and  $f$  is the frequency [16]. The two time constants describe the equalization curve, but in some cases,  $t_1$  is  $\infty$ . For the 7.5 ips audio tape recording,  $t_1$  and  $t_2$  are, respectively,  $\infty$  and  $70 \mu\text{s}$  for CCIR but  $3180 \mu\text{s}$  and  $50 \mu\text{s}$  for NAB (see Table 1). The characteristics of these equalization standards will be analyzed in this paper. Figures 1(a) and 1(b) present the frequency response of pre- and postemphasis curves, respectively, for NAB and CCIR equalization. An incorrect juxtaposition of the pre- and postemphasis significantly alters the spectrum and therefore requires compensation to avoid the loss of accuracy for digitized audio documents. Starting from these analytic formulas, the paper will describe how to create digital filters of the pre- and postemphasis curves to digitally compensate equalization errors in digitized 7.5 ips audio tape recordings.

### 3. Digital Equalizations

**3.1. Signals and a Chain of Filters.** Given an analogue signal  $x \in L^2(\mathbb{R})$ , it passes through two steps before digitization: a recording phase and a reproducing phase. An equalization for each step is defined, followed by the convolutions of the signals with the impulse responses of the recording and reproducing filters  $w \in L^2(\mathbb{R})$  and  $r \in L^2(\mathbb{R})$ , denoted, respectively, with  $\mathcal{W}(x) = x * w$  and  $\mathcal{R}(x) = x * r$ . The resulting filter is defined as  $\mathcal{E}(x) = \mathcal{R} \circ \mathcal{W}(x) = (x * w) * r$ .

Considering the transfer functions of our filters, in this context, a correct equalization  $\mathcal{E}: L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$  of a signal has to be a flat equalization, which means that its transfer function is the identity operator, i.e.,  $E = \text{id}$ , where  $E: \mathbb{C} \rightarrow \mathbb{C}$  is the transfer function of  $\mathcal{E}$ . Denoting, respectively, with  $W$  and  $R: \mathbb{C} \rightarrow \mathbb{C}$  the transfer functions of the recording and reproducing filters  $\mathcal{W}$  and  $\mathcal{R}$ , we should have  $R = W^{-1}$ . In this project, however, we are dealing with a nonflat equalization  $\tilde{\mathcal{E}} = \tilde{\mathcal{R}} \circ \mathcal{W}$ , where its transfer function  $\tilde{E} = W \cdot \tilde{R} \neq \text{id}$  since the reproducing curve  $\tilde{R}$  is wrongly set. It is necessary to apply a filter  $\mathcal{F}: L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$  in order to obtain a flat equalization:  $\tilde{E} \cdot F = \text{id}$ , where  $F: \mathbb{C} \rightarrow \mathbb{C}$  is the transfer function of  $\mathcal{F}$ .

Taking advantage of the structure of  $\tilde{\mathcal{E}}$ , it is possible to express the desired transfer function  $F$  as

$$F = \tilde{R}^{-1} \cdot W^{-1}. \quad (3)$$

This last equality is a solution in terms of the transfer functions obtained from the standard NAB and CCIR equalizations defined in [15, 16].

**3.2. Standard NAB and CCIR Transfer Functions.** From the standard references, the reproducing characteristic curves are given by magnitude function (2). By definition, (2) is derived from the transfer function of the reproducing analogue filter. Since the standards consider only first-order low-pass and high-pass filters, it is possible to show that

$$R(s) = \frac{st_1(1 + st_2)}{1 + st_1}, \quad (4)$$

where  $s \in \mathbb{C}$  is the transfer function needed. Computing the squared norm of (4) on an imaginary line  $i\omega$ , where  $\omega \in \mathbb{R}$ , the result is

$$|R(i\omega)|^2 = R(i\omega)\overline{R(i\omega)} = \frac{|i\omega t_1(1 + i\omega t_2)|^2}{|1 + i\omega t_1|^2} = \frac{\omega^2 t_1^2(1 + \omega^2 t_2^2)}{1 + \omega^2 t_1^2}, \quad (5)$$

which is the squared argument in (2).

The transfer function  $W$  is a rational complex polynomial given by the inverse of  $R$ :

$$W(s) = R^{-1}(s) = \frac{1 + st_1}{st_1(1 + st_2)}, \quad (6)$$

where  $s \in \mathbb{C}$ .

Since, in our case, we can infer that the parameters of  $\tilde{R}$  are incorrect, equation (3) becomes

$$F(s) = \frac{t_1(1 + st_2)(1 + st_3)}{t_3(1 + st_4)(1 + st_1)}, \quad (7)$$

where  $s \in \mathbb{C}$ ,  $t_1, t_2$  are the parameters of the recording transfer function  $W$ , and  $t_3, t_4$  are the parameters of the wrongly reproduced transfer function  $\tilde{R}$ .

**3.3. Filter Stability.** Now that the general structure of the corrective filters has been described, it is necessary to verify that, with all possible combinations of the four parameters  $t_1, t_2, t_3$ , and  $t_4$ , stable filters are obtained. From the reference tables for standard NAB and CCIR equalizations [15, 16], coefficients  $t_1$  and  $t_3$  can assume finite values or can be  $\infty$ . This means that, considering (7) as a function with parameters  $t_1$  and  $t_3$ , there are four cases:

- (i)  $t_1, t_3 < \infty$ : no change in the formal structure of (7)
- (ii)  $t_1, t_3 = \infty$ : (7) becomes  $\lim_{t_1, t_3 \rightarrow \infty} F(s) = ((1 + st_2)/(1 + st_4))$
- (iii)  $t_1 < \infty$  and  $t_3 = \infty$ : (7) becomes  $\lim_{t_3 \rightarrow \infty} F(s) = ((t_1(1 + st_2))/((1 + st_4)(1 + st_1)))$
- (iv)  $t_1 = \infty$  and  $t_3 < \infty$ : similarly,  $\lim_{t_1 \rightarrow \infty} F(s) = (((1 + st_2)(1 + st_3))/(t_3(1 + st_4)))$

Also, all these filters except the last are stable as they have poles when  $s = -(1/t_1)$  or  $s = -(1/t_4)$ , which are both strictly negative. The fourth case gives an unstable filter with the pole in  $s = 0$ .

Clearly, the real case which corresponds to the unstable filter is relevant in applications as it is the inverse of the chain  $F = \tilde{R}^{-1} \cdot W^{-1}$  (i.e.,  $F^{-1} = \tilde{R} \cdot W$ ), where  $W$  and  $\tilde{R}$  are, respectively, the transfer functions of CCIR and NAB equalizations (see Table 2 as a summary of all cases).

We need to approximate the unstable filter with a stable one, which is sufficiently "close" (clarified in the following section) to the first, to produce a similar equalization. Formally, we digitize this filter via bilinear transform and

TABLE 1: Time constants of the CCIR and NAB postemphasis curves at 7.5 ips.

Equalization standard	$t_1$ ( $\mu\text{s}$ )	$t_2$ ( $\mu\text{s}$ )
CCIR	$\infty$	70
NAB	3180	50

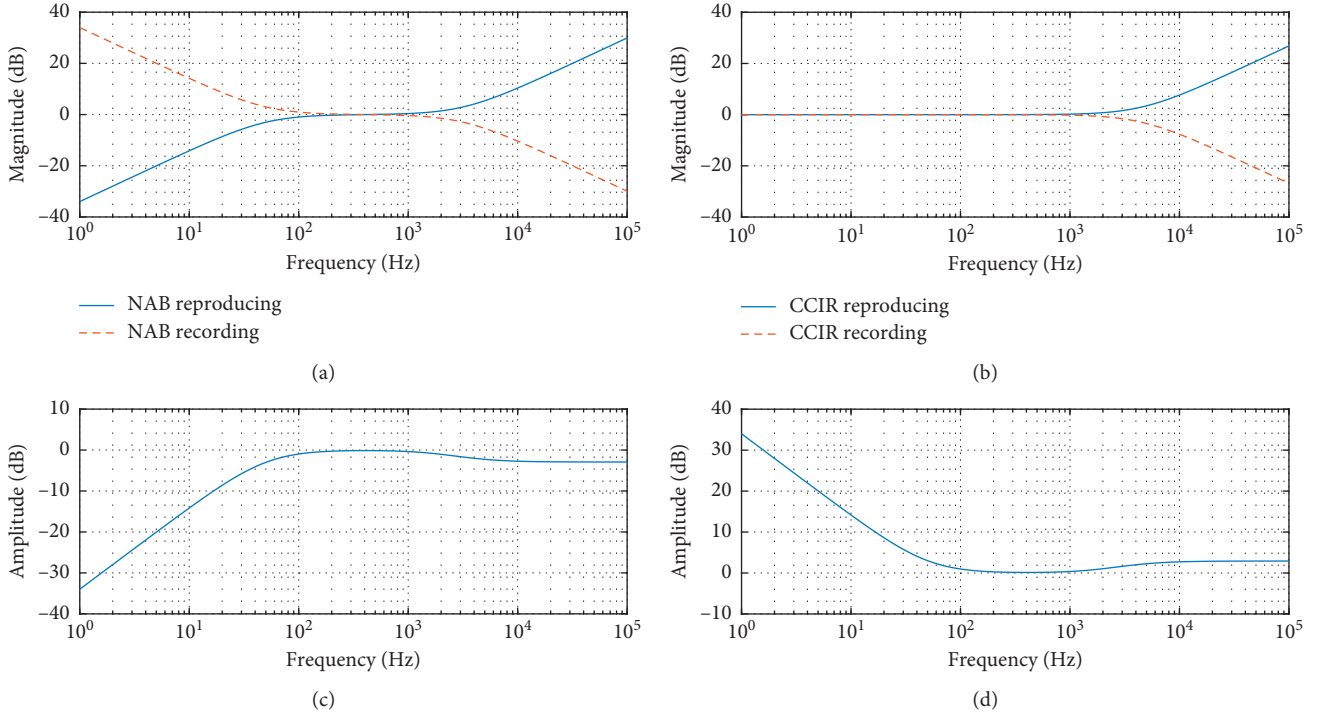


FIGURE 1: Plots of the frequency response of NAB and CCIR pre- and postemphasis curves, respectively (a, b) and the filters for compensating the incorrect juxtapositions: NAB-CCIR (c) and CCIR-NAB (d). (a) NAB equalization. (b) CCIR equalization. (c) Filter for the NAB-CCIR correction. (d) Filter for the CCIR-NAB correction.

TABLE 2: Time constants of the considered cases: in the stable one,  $W$  (parameters  $t_1$  and  $t_2$ ) used NAB values, but  $\tilde{R}$  (parameters  $t_3$  and  $t_4$ ) used CCIR values; in the unstable one,  $W$  used CCIR values, but  $\tilde{R}$  used NAB values.

Time constant	Stable ( $\mu\text{s}$ )	Unstable ( $\mu\text{s}$ )
$t_1$	3180	$\infty$
$t_2$	50	70
$t_3$	$\infty$	3180
$t_4$	70	50

digitally approximate it by solving a minimum least square problem, as explained in the following.

**3.4. Digital Approximation of the Unstable Filter.** After the digitization of the transfer function  $F(s)$ , the MATLAB function “freqz” was used to study the behavior of the unstable filter. Examining its output, it was observed that the frequency vector reaches  $\infty$  in its first cell, near 0 Hz as expected. Using a pragmatic approach, this value has been overridden with 0. Since this is an extreme modification of the frequency response vector, it has been studied if it is possible to find an approximated stable transfer function starting from the modified frequency response vector such

that its frequency response is close to the analogue transfer function, at the very least in audible frequencies.

The transfer function we are dealing with is a rational function  $H: \mathbb{C} \rightarrow \mathbb{C}$  of the form

$$H(z) = \frac{B(z)}{A(z)} = \frac{b(0)z^m + b(1)z^{m-1} + \dots + b(m-1)z + b(m)}{a(0)z^n + a(1)z^{n-1} + \dots + a(n-1)z + a(n)}, \quad (8)$$

where  $z \in \mathbb{C}$  and  $B, A: \mathbb{C} \rightarrow \mathbb{C}$  are complex polynomials of finite degrees  $m, n \in \mathbb{N}$ , with coefficient vectors  $b \in \mathbb{C}^{m+1}$  and  $a \in \mathbb{C}^{n+1}$ , respectively. Given a vector of frequency points  $f \in \mathbb{R}^l$ , where  $l \in \mathbb{N}$ , and the corresponding frequency response vector  $h \in \mathbb{C}^l$ , we define the following minimization problem:

$$\min_{a,b} \sum_{k=1}^l \left| h(k) - \frac{B(f(k))}{A(f(k))} \right|^2. \quad (9)$$

A solution to this problem, given via an algorithm based on a damped Gauss–Newton iterative search method—described in [17] and implemented in the MATLAB function “invfreqz”, is the coefficient vectors of the stable rational transfer function approximating the unstable filter we found in the previous section. The inputs of this function are the frequency vector  $f$ , the frequency response vector  $h$  (modified as described at the beginning of this section), the polynomial degrees  $m$  and  $n$  of the numerator and denominator of the solution, and the number of iterations “iter.” We set  $m$  and  $n$  equal to 2 in order to maintain the same general structure of this kind of filter, and we set “iter” to 10 since at this point, the approximation converges.

In Figure 2(a), the “bilinear approximated” curve is obtained by using the MATLAB function “freqz” applied to the approximated stable transfer function, i.e., the output of “invfreqz.” Figures 2(b) and 2(c) quantify the approximation in further detail, specifically focusing on low and high frequencies, respectively. At low frequencies, it is noticeable that the resolution of the “bilinear” curve is poor. However, this study primarily aims to characterize the feasibility of our earlier noted pragmatic approach: subsequent studies for improving approximation could be done in the future.

The described approximation method satisfies the stability requirement and produces a transfer function with frequency response close to the original. Given the nature of approximations, future investigations could lead to different solutions; for example, it is possible to modify the analogue transfer function by adding a pole centered at a very low frequency.

#### 4. Assessment of Perception

We conducted an experiment with the aim of assessing the perception of similarity for various equalization curves applied to the same stimulus. We adopted an approach inspired by the MULTiple Stimuli with Hidden Reference and Anchor (MUSHRA) test, a well-established method for evaluating the quality of several versions of an audio stimulus [18, 19]. Our MUSHRA-inspired assessment aimed to investigate whether or not musically trained participants were able to distinguish a stimulus recorded in the magnetic tape that is digitized with a correct equalization standard (Reference) from (a) the same stimulus that is digitized with an intentionally incorrect equalization standard (Foil) and (b) the incorrect stimulus that has been subsequently corrected with the digital filters proposed in the previous section. For (b), two separate correction versions are proposed. In the first version, the incorrect stimulus was directly corrected with a MATLAB script, whereas in the second version, an *ad hoc* web interface adopting the Web Audio API was used to correct the stimulus in order to simulate the use of the filters in web tools for accessing historical audio documents, such as in [13] (see more details in Section 4.3).

**4.1. Materials.** The experiment contained 8 audio stimuli, listed in Table 3. As will be detailed in Procedures, 6 stimuli were used for assessment, and 2 stimuli were used as training. Each stimulus was a 10-second excerpt of an electroacoustic composition, chosen from important repertoire of the genre. The stimuli were selected to produce a heterogeneous set from a spectral standpoint, with each exhibiting a wide range of frequency combinations and textures. Additionally, half of the stimuli were produced with a NAB preemphasis curve and the other half with a CCIR preemphasis curve.

For each stimulus, there were 6 different equalizations (henceforth filters) provided. See Procedures for production details of each filter. The 6 filters were as follows:

- (1) “Reference”: the correctly produced equalization standard.
- (2) “Hidden reference”: an exact copy of the “Reference” audio, used as an accuracy check.
- (3) “Anchor”: the Reference processed with a low-pass filter. This was easily discernible from the other filters and so was used as a second accuracy check.
- (4) “Foil”: an intentionally incorrect equalization, created by mismatching the recording and reproducing curves.
- (5) “MATLAB correction”: a subsequent correction of the Foil audio using a MATLAB script.
- (6) “Web Audio API correction”: a subsequent correction of the Foil audio using an *ad hoc* web interface.

**4.2. Participants.** Twenty-three participants were recruited from an undergraduate music course in Australia. Thirteen participants (57%) were male, and 10 (43%) were female. Participants ranged in age from 18 to 33 years ( $M = 19.7$ ;  $SD = 3.7$ ) and were asked how many years they had received music training (range: 1–20,  $M = 11.1$ , and  $SD = 3.7$ ; all but 1 participant reported 7 or more years of music training). Prospective participants all agreed to participate and completed a written consent form. The study received ethics approval (UNSW Human Ethics Approval HC13015).

**4.3. Procedures.** Participants were tested in groups  $\leq 5$ . Testing was conducted on MacBook Pro laptops (13 inches, mid-2010) with Sennheiser HD280 Pro headphones. The web interface of the test was created by using BeagleJS, a framework based on HTML5 and JavaScript [20], and the browser Google Chrome was used for all tests. The loudness was set consistently on each laptop, and the loudness toggle keys were locked for each computer. The loudness level was inspected by the research team by measuring the sound pressure level for a pink noise sound file using a Testo 815 meter. Measurements were taken with the following setting: slow time weighting, “A” frequency weighting, a measurement range of 50 to 100 dB, and a “maximum” hold function. Ten measurements were made on each laptop, switching to a second pair of headphones after the fifth

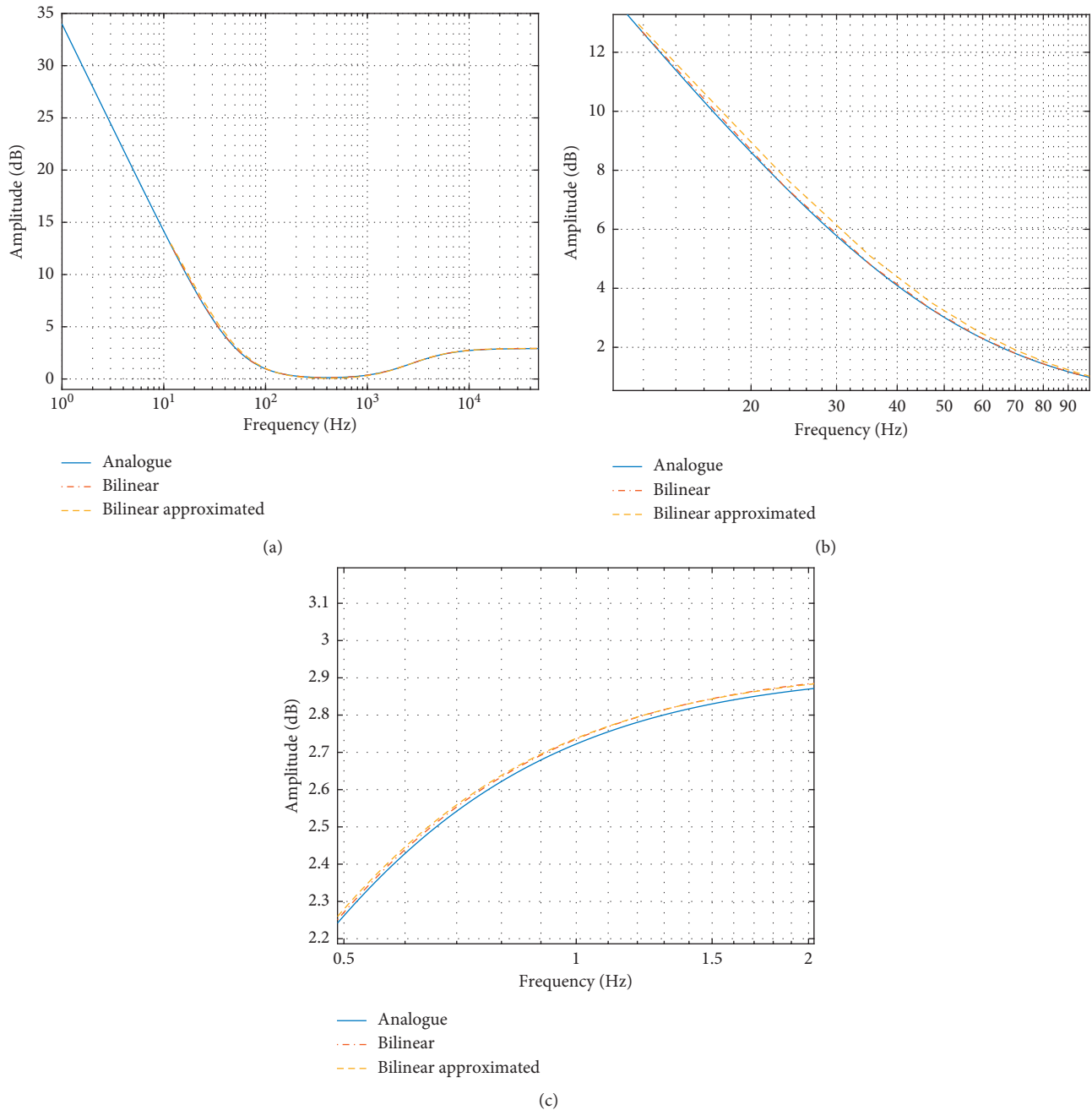


FIGURE 2: (a) A comparison between the frequency response of the analog unstable filter (“analogue”), the digitized unstable filter (“bilinear”), and the approximated stable filter (“bilinear approximated”). (b, c) Approximation quantification in further detail, specifically focusing on low and high frequencies, respectively. (a) Overall comparison. (b) Low-frequency approximation. (c) High-frequency approximation.

measurement. Measurements ranged from 78.2 dB to 81.2 dB across all laptops, with  $M = 80.5$  and  $SD = 0.9$ .

The experiment consisted of 8 different tests, with each concerning one of the 8 stimuli in Table 3. Each test was presented on a single screen of the interface (see Figure 3) and contained all 6 filters for the examined stimulus—Reference, Hidden reference, Anchor, Foil filter (CN foil or NC foil), MATLAB correction, and Web Audio API correction. As per the MUSHRA protocol [19], the Reference filter was always the first filter presented and was clearly

labeled, whereas the remaining filters were randomized and unlabeled.

Participants were able to replay each audio file as often as they wished and in any order and were tasked with evaluating the similarity of each of the unlabeled filters in comparison to the Reference. Responses were recorded on 11-point rating scales (0–10) corresponding to “different”; “somewhat different”; “slightly different”; “nearly identical”; and “identical” (see Figure 3). Furthermore, in line with the MUSHRA guidelines [19], participants received a “training

TABLE 3: The 8 audio stimuli used for the training and test examples.

Stimulus	Preemphasis	Phase
Edgard Varèse <i>Poème électronique</i>	CCIR	Training
György Ligeti <i>Artikulation</i>	CCIR	Test
Bruno Maderna <i>Musica su due dimensioni</i>	CCIR	Test
Luciano Berio <i>Différences</i>	CCIR	Test
Jonathan Harvey <i>Mortuos Plango, Vivos Voco</i>	NAB	Training
Luciano Berio <i>Visage</i>	NAB	Test
Bruno Maderna <i>Syntaxis</i>	NAB	Test
Bruno Maderna <i>Continuo</i>	NAB	Test

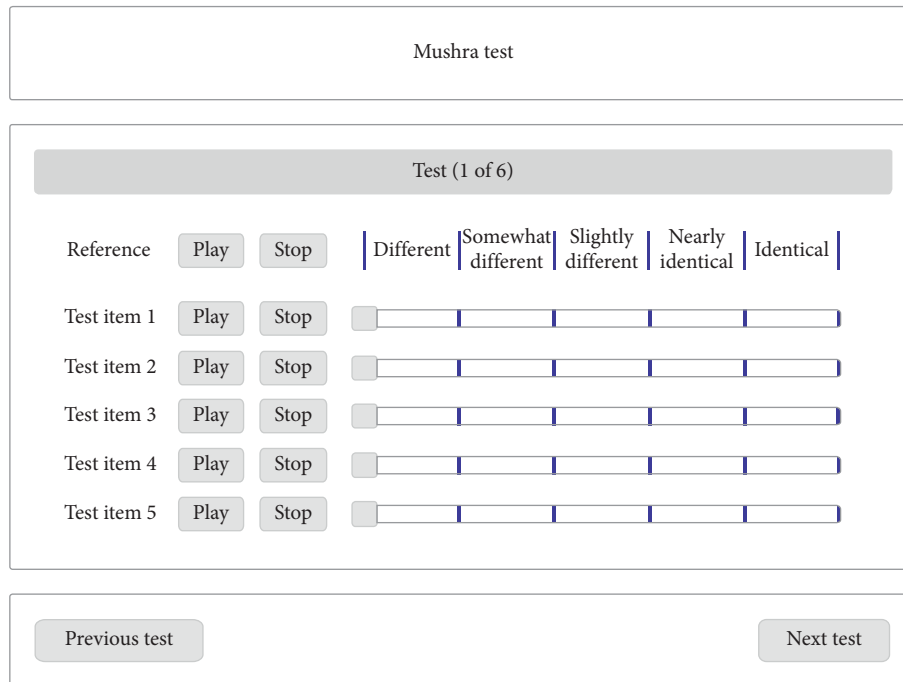


FIGURE 3: Screenshot of the MUSHRA-inspired test interface, showing one of the 6 nontest examples.

phase” for the first two stimuli (*Poème électronique* and *Mortuos Plango, Vivos Voco*). For the training phase, all filters were labeled, and ratings were not recorded; thus, we refer to only 6 test stimuli in the analysis.

To create the 6 filters, high-quality digital samples from a computer were recorded onto a new tape using the professional Studer A810 with a recording speed of 7.5 ips and the CCIR preemphasis curve for the first four stimuli and NAB for the second four. After this stage, the Reference filter was obtained for each stimulus through the digitization of the recorded samples with the correct juxtaposition of the inverse analogue filter used during the recording. A second version of the Reference (Hidden Reference) was also included in the test phase. Starting from the Reference, the Anchor filter was obtained for each stimulus by processing the Reference with a low-pass filter measuring  $-3$  dB at 3.5 kHz as defined by the MUSHRA standard [19]. Next, the signals that were recorded onto the tape for the creation of the Reference were digitized a second time, using an uncorrected inverse analogue filter, i.e., CCIR as the preemphasis curve and NAB as the postemphasis curve (CN

foil), and vice versa (NC foil). The Foil digitization is used to simulate the real-life situation where the incorrect post-emphasis curves are selected.

Finally, each Foil filter was compensated, from the spectral point of view, with the two correction filters described in Section 4.1: a MATLAB script and an *ad hoc* web app based on the Web Audio API. The version obtained by the MATLAB correction implements a high-resolution offline processing of the signal, whereas the version obtained with the web app performs a real-time processing of the signal with ConvolverNode [21]. This node convolves the audio signal with an impulse response of the filters, and the resulting signal was recorded using professional equipment and normalized.

**4.4. Preliminary Analysis.** As per the MUSHRA guidelines [19], the Hidden reference and the Anchor filter each acted as a reliability test. Any participants that rated the Hidden reference  $\leq 7$  (constituting the rating of “nearly identical” or lower) were removed from the entire data sample. Similarly,

TABLE 4: Descriptive statistics of similarity ratings by stimulus and filter type for the reduced participant sample size ( $n = 14$ ). Results are displayed as mean (standard deviation).

Stimulus	Hidden	MATLAB	Foil	Anchor	Web
Visage	8.36 (1.15)	7.57 (2.14)	7.64 (1.98)	0.50 (0.52)	7.64 (1.45)
Syntaxis	8.07 (0.83)	7.14 (1.51)	6.43 (2.38)	0.64 (0.50)	5.71 (2.37)
Continuo	8.21 (1.12)	6.71 (2.20)	5.43 (2.50)	0.57 (0.51)	4.64 (2.17)
Artikulation	8.07 (1.00)	6.64 (2.13)	7.14 (1.41)	0.57 (0.51)	7.21 (1.85)
Differences	8.00 (1.11)	7.50 (2.17)	6.57 (2.59)	0.50 (0.52)	7.36 (1.91)
Musica su due dimensioni	7.50 (0.76)	6.93 (2.27)	6.43 (1.95)	1.21 (1.12)	6.00 (2.51)

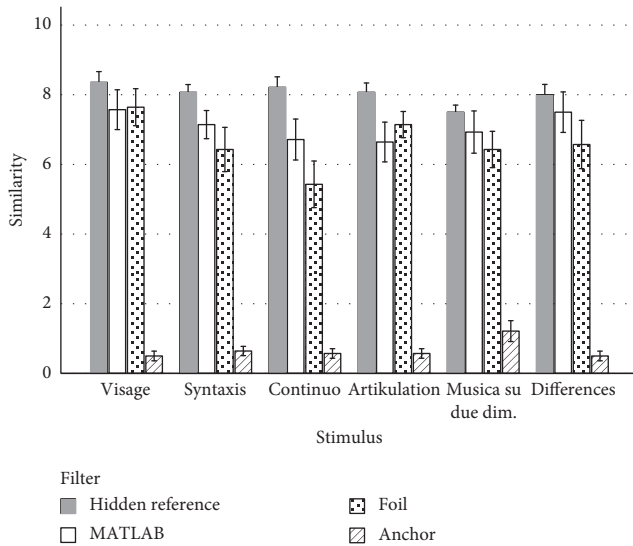


FIGURE 4: Mean similarity ratings by stimulus and filter ( $n = 14$ ). Error bars are equal to  $\pm 1$  SE.

any participants who rated the Anchor  $\geq 3$  (constituting a rating of “somewhat different” or higher) were removed from the data sample. This produced a subsample of  $n = 14$  reliable participants.

Before analysis of results, we listened to each filter for all stimuli. This listening examination suggested that, in all cases, the Web Audio API correction was accompanied by an unintentional, perceivable equalization effect. To further investigate this, long-term average spectrum (LTAS) plots on each filter for all stimuli were calculated. The LTAS plot of the Web Audio API correction filter for each stimulus was visually different to that of the Reference filter plot, whereas the MATLAB correction filter plot was not. Therefore, we identified a production error in the method used to create the Web Audio API correction filter and so exclude this filter from all subsequent analyses (although for interest, we retain descriptive statistics for the Web Audio API filter in Table 4).

**4.5. Results and Discussion.** Two separate two-way within-subject ANOVAs were performed, with the first examining the three NAB stimuli and the second examining the three CCIR stimuli. The ANOVAs were used to investigate any differences in similarity (dependent variable), with filter and stimulus each used as within-subject independent variables. Descriptive statistics for each stimulus and filter are reported

in Table 4 and are plotted in Figure 4. Anchor filters were not included in ANOVA analyses because of the following: (1) this investigation is concerned with interactions between the Reference, Foil, and correction filters, whereas Anchor filters are designed to examine reliability (see previous section); (2) the inclusion of these data, which are consistently rated lower in similarity in comparison to the remaining filters, would likely violate the assumption of normality [22]. Regardless, we included descriptive statistics for the Anchor filters in Table 4 and performed separate paired sample  $t$ -tests between ratings for the Hidden reference and the Anchor filters (included in Table 5). As similarity ratings for this filter were consistently lower than those for the MATLAB and Foil filters, these data confirm that participants were able to perceive the effects of the production error.

The two ANOVAs each produced a significant main effect of the filter: NAB ( $F(2, 26) = 10.54, p < 0.001, \eta^2 = 0.448$ ) and CCIR ( $F(2, 26) = 5.47, p = 0.010, \eta^2 = 0.296$ ). There were no significant interactions between the independent variables for either of the ANOVAs. We performed two types of post hoc analysis. First, for each stimulus, we examined differences in similarity ratings between the Hidden reference and either the MATLAB or Foil filter; these results are reported in Table 6. However, due to the small sample size ( $n = 14$ ), this may not produce sufficient statistical power for meaningful analysis [23]. Therefore, we also compared similarity ratings for the Hidden reference filter with the MATLAB and Foil filters, collapsed either across the 3 NAB stimuli or across the 3 CCIR stimuli (see Table 5).

It is evident from the data reported in Tables 4 and 5 that participants were able to distinguish between the Anchor filters and all remaining filters, regardless of the stimulus examined. Post hoc results reported in Table 6 (approach 1) suggest that the MATLAB correction was not perceivable from the Hidden reference for 5 of the 6 stimuli (all except *Artikulation*). In contrast, the Foil filters were rated statistically lower in similarity than the Hidden reference filter for 3 of the 6 stimuli and approached significance ( $p = 0.088$ ) for 2 of the remaining 3 stimuli. For the stimulus *Visage*, similarity ratings for the MATLAB and Foil filters occur in close proximity to each other. As this anomalous result occurs only for this stimulus, we suggest that it may be the result of complex music textures within this composition, which could have created difficulties in differentiating between the filters. For approach 2, where the data were collapsed prior to post hoc testing (either as NAB or CCIR;



TABLE 5: Results of paired sample  $t$ -tests between the Hidden reference and the remaining filters. Tests have been collapsed across the 3 NAB stimuli or the 3 CCIR stimuli. Bonferroni correction has been applied to  $p$  values. Effect size refers to Cohen’s  $d$ .

Filter comparison	Filter type	Significance ( $p$ )	Effect size ( $d$ )
MATLAB	NAB	0.199	1.04
NC foil	NAB	0.006	2.24
Anchor	NAB	<0.001	1.81
MATLAB	CCIR	0.141	0.79
CN foil	CCIR	0.019	1.04
Anchor	CCIR	<0.001	1.19

TABLE 6: Results of paired sample  $t$ -tests between the Hidden reference and the remaining filters by stimulus. Bonferroni correction has been applied to  $p$  values. Effect size refers to Cohen’s  $d$ .

Stimulus	Compared filter	Significance ( $p$ )	Effect size ( $d$ )
Visage	MATLAB	0.538	0.46
	NC foil	0.348	0.44
Syntaxis	MATLAB	0.054	0.72
	NC foil	0.020	0.95
Continuo	MATLAB	0.082	0.86
	NC foil	0.010	1.43
Artikulation	MATLAB	0.042	0.86
	CN foil	0.034	0.76
Differences	MATLAB	0.340	0.46
	CN foil	0.088	0.66
Musica su due dimensioni	MATLAB	0.316	0.47
	CN foil	0.088	0.61

see Table 5), no significant differences were observed between the Hidden reference and MATLAB filters for either NAB or CCIR stimuli. As significant differences were observed between the Hidden reference and Foil filters for both stimulus types (NAB and CCIR), we conclude that overall, the MATLAB correction appears to be a successful method for compensating existing digitization errors. Furthermore, it is important to note that the Hidden reference was consistently rated lower than the maximum similarity level of 10, despite the audio file being identical to the original Reference file. This result suggests the presence of a rating bias [24] in which participants appear hesitant to use the extreme ends of the rating scale.

## 5. Conclusion

This paper investigated the equalization problem for the active preservation process of audio tape recordings. Proper selection of equalization in the digitization process is essential in preserving the historical authenticity of an audio work, although the differences between the original (“correct”) and arbitrary equalizations may be subtle to an untrained listener. We investigated tools to compensate equalization errors introduced in the re-recording process. With these tools, an archivist or musicologist who notices an error in the preservation master (through listening or with automatic tools [3, 25]) can make a correction and so provide an authentic listening experience without having to recover the original analogue audio document or perform redigitization. A MUSHRA-inspired test was conducted on six electroacoustic stimuli to investigate perceivable

differences between (a) correctly digitized “Reference” versions, (b) two intentionally incorrect “Foil” versions (in terms of the digitization process), (c) easily distinguishable 3.5 kHz “Anchor” filters, and (d) subsequent digital correction filters of the Foil equalizations. Two digital filters were initially presented to compensate equalization errors in the case of 7.5 ips recordings both with NAB and CCIR standards, although one of these correction filters (the Web Audio API correction filter) contained a production error and so had to be removed from analyses. Therefore, the present study was only able to evaluate the validity of the MATLAB correction filter.

Similarity ratings were examined with two ANOVAs, and two distinct post hoc approaches were taken. When data were collapsed either as NAB or CCIR stimuli (prior to post hoc testing), participants were not able to distinguish between the Hidden reference filter and the MATLAB correction filter. In comparison, both types of Foil filters and the Anchor filter produced significantly lower ratings of similarity than the Hidden reference. As such, we conclude that the MATLAB correction filter is a promising method to aid in the preservation of analogue works.

Five design issues were identified. First, future studies should include a larger sample size and aim to incorporate historically informed expert listeners who are highly familiar with and knowledgeable about electroacoustic music. Such an inclusion should increase reliability in comparison to the undergraduate music students who were used in the present study. Second, comparisons with additional correction filters (as was originally intended in this design) would allow further clarification on the accuracy of the MATLAB

correction. Third, one of the stimuli in this study (Visage) produced an anomalous result in which ratings for the MATLAB and Foil filters occurred within very close proximity to each other. We suggest that this may be a result of complex music textures within the composition, which could produce difficulty in differentiating between filters. This result highlights the need for future designs to place great care on stimulus selection. Fourth, the results in this study suggest the presence of a rating scale bias in which participants are hesitant to use the extreme ends of the rating scale. Additional rating biases may also be present, such as the range equalizing bias [24]. Specifically, the differences between the clearly discernible Anchor filter compared to the remaining, less-discernible filters might produce a response in which differences between the less-different filters become comparatively difficult to perceive. Thus, we recommend that future studies adopt a between-subjects design that investigates the impact of Anchor filters on ratings of the remaining filters, such as through a “MUSHRA versus MUSHR” test (with the latter containing no Anchor filters). Finally, while it is beyond the scope of the present paper, future studies could expand this research area by examining additional corrective equalization methods for other equalization standards (that is, other than NAB and CCIR) and at playback speeds other than 7.5 ips. However, in such a case, numerous factors must be considered, such as the changes in curves between equalization standards at various playback speeds, as well as the effect on the frequency response of the filters derived from the change of speed.

The preservation and ongoing authentic use of historical audio documents hinges on the application of multimedia information processing tools, with particular attention on the parameters that were used at the time of the recording, as well as their metadata. The tools presented in this paper are aimed to produce a complete and historically informed use of historical audio (words, sound effects, and music) for multimedia archives.

### Data Availability

The ethics approval for this research does not allow for the release of the experimental data used to support these findings, even if anonymized.

### Conflicts of Interest

The authors declare no conflicts of interest related to this work.

### Acknowledgments

The authors wish to acknowledge Roberto Rinaldo (University of Udine, Italy) and Roberto Barumerli (University of Padova, Italy) for helpful discussions and suggestions. This submission was supported in part by an Australian Research Council Future Fellowship (FT120100053) held by Emery Schubert.

### References

- [1] F. Bressan, A. Rodà, S. Canazza, F. Fontana, and R. Bertani, “The safeguard of audio collections: a computer science based approach to quality control – the case of the sound archive of the arena di Verona,” *Advances in Multimedia*, vol. 2013, p. 14, Article ID 276354, 2013.
- [2] 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, Article ID 489515, 21 pages, 2013.
- [3] 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, no. 2, 2018.
- [4] K. Bradley, *IASA TC-04 Guidelines in the Production and Preservation of Digital Audio Objects: Standards, Recommended Practices, and Strategies*, International Association of Sound and Audio Visual Archives, London, UK, 2nd edition, 2009.
- [5] V. Valimaki and J. D. Reiss, “All about audio equalization: solutions and frontiers,” *Applied Sciences*, vol. 6, p. 5, 2016.
- [6] J. C. Mallinson, “Tutorial review of magnetic recording,” *Proceedings of the IEEE*, vol. 64, no. 2, pp. 196–208, 1976.
- [7] L. D. Fielder, “Pre-and postemphasis techniques as applied to audio recording systems,” *Journal of the Audio Engineering Society*, vol. 33, no. 9, pp. 649–658, 1985.
- [8] M. Camras, *Magnetic Recording Handbook*, Van Nostrand Reinhold Co., New York, NY, USA, 1988.
- [9] P. Copeland, *Manual of Analogue Sound Restoration Techniques*, British Library, London, UK, 2008.
- [10] 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)*, pp. 17–24, Bari, Italy, January 2017.
- [11] D. Schüller, “The ethics of preservation, restoration, and Reissues of historical sound recordings,” *Journal of Audio Engineering Society*, vol. 39, pp. 1014–1016, 1991.
- [12] S. Canazza, C. Fantozzi, and N. Pretto, “Accessing tape music documents on mobile devices,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 12, no. 1s, pp. 1–20, 2015.
- [13] C. Fantozzi, F. Bressan, N. Pretto, and S. Canazza, “Tape music archives: from preservation to access,” *International Journal on Digital Libraries*, vol. 18, no. 3, pp. 233–249, 2017.
- [14] N. Pretto, A. Russo, Federica 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, SMC20*, pp. 394–398, Torino, Italy, June 2020.
- [15] 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*, IEC, Geneva, Switzerland, 1994.
- [16] NAB, *Magnetic Tape Recording and Reproducing (Reel-To-Reel)*, NAB, Melbourne, Australia, 1965.
- [17] J. E. Dennis Jr and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM, Philadelphia, PA, USA, 1996.
- [18] D. Marston and A. Mason, “Cascaded audio coding,” *EBU Technical Review*, vol. 304, 2005.

- [19] International Telecommunications Union, *BS. 1534-3. Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems. Recommendation*, ITU-R, Geneva, Switzerland, 2015.
- [20] S. Kraft and U. Zölzer, “BeaqlJS: HTML5 and JavaScript based framework for the subjective evaluation of audio quality,” in *Proceedings of the Linux Audio Conference*, Karlsruhe, DE, USA, May 2014.
- [21] P. Adenot and C. Wilson, *Web Audio API (W3C Working Draft)*, Springer, Berlin, Germany, 2015.
- [22] B. Lantz, “The impact of sample non-normality on ANOVA and alternative methods,” *British Journal of Mathematical and Statistical Psychology*, vol. 66, no. 2, pp. 224–244, 2013.
- [23] C. R. W. VanVoorhis and B. L. Morgan, “Understanding power and rules of thumb for determining sample sizes,” *Tutorials in Quantitative Methods for Psychology*, vol. 3, no. 2, pp. 43–50, 2007.
- [24] S. Zielinski, P. Hardisty, C. Hummersone, and F. Rumsey, “Potential biases in MUSHRA listening tests,” *Audio Engineering Society Convention*, vol. 123, 2007.
- [25] 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.