A workflow and novel digital filters for compensating speed and equalization errors on digitized audio open-reel tapes

Niccolò Pretto
niccolo.pretto@dei.unipd.it
Department of Information Engineering, University of Padova
Padova, IT

Nadir Dalla Pozza
dallapozza@dei.unipd.it
Department of Information Engineering, University of Padova
Padova, IT

Alberto Padoan
padoanalbe@dei.unipd.it
Department of Information Engineering, University of Padova
Padova, IT

Anthony Chmiel
a.chmiel@westernsydney.edu.au
Western Sydney University
Sydney, AU

Kurt James Werner
kurt.james.werner@gmail.com
iZotope, Inc
Cambridge, Massachusetts, US

Alessandra Micalizzi
a.micalizzi@sae.edu
SAE Institute
Milan, IT

Emery Schubert
e.schubert@unsw.edu.au
Empirical Musicology Laboratory, University of New South Wales
Sydney, AU

Antonio Rodà
antonio.roda@dei.unipd.it
Department of Information Engineering, University of Padova
Padova, IT

Simone Milani
simone.milani@dei.unipd.it
Department of Information Engineering, University of Padova
Padova, IT

Sergio Canazza
sergio.canazza@dei.unipd.it
Department of Information Engineering, University of Padova
Padova, IT

ABSTRACT

This paper presents a workflow and novel digital filters for compensating speed and equalization errors that can impact digitized audio open-reel tapes. We examine three frequent cases of mismatch between recording and reproducing standards: NAB 3.75 ips - CCIR 7.5 ips; NAB 3.75 ips - CCIR 15 ips; NAB 7.5 ips - CCIR 15 ips. Three MUSHRA-inspired tests ("sets") containing ≥21 participants were used to perceptually assess the workflow and digital filters, using excerpts of music and voice. The results indicated that the digital correction filters performed well, although the electroacoustic stimuli in Set C provided mixed results, suggesting that the style of the music used in perception tests should not be overlooked.

CCS CONCEPTS

• **Applied computing** → **Sound and music computing**; • **Hardware** → **Digital signal processing**.

KEYWORDS

equalization, digital filters, analog recordings, open-reel tape

Audio Mostly '21, September 1–3, 2021, Trento, Italy

ACM ISBN 978-1-4503-8569-5/21/09.
https://doi.org/10.1145/3478384.3478409

ACM Reference Format:
Niccolò Pretto, Nadir Dalla Pozza, Alberto Padoan, Anthony Chmiel, Kurt James Werner, Alessandra Micalizzi, Emery Schubert, Antonio Rodà, Simone Milani, and Sergio Canazza. 2021. A workflow and novel digital filters for compensating speed and equalization errors on digitized audio open-reel tapes. In *Audio Mostly*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3478384.3478409

1 INTRODUCTION

Audio recordings constitute an important part of cultural heritage and a priceless source of information for several research areas such as linguistics, anthropology, and musicology. Data transfer onto new media (re-recording) is essential for preventing an irreversible partial or complete loss of information due to the degradation of the original signal [20]. Analog recordings require a digitization process, although this process is not neutral. It can introduce artifacts, and furthermore aspects concerning the reproduction of the original source need to be considered from a philological point of view, particularly with regard to breaches in authenticity [8]. In recent decades, the international community has placed considerable effort in digitization, often with massive digitization projects. In some cases, the digitization tasks were performed without auditory supervision. This can lead to digitization errors, which are sometimes not identified until months or years after the task. If the error is detected after the digitization project, it may not be possible to perform a new digitization due to lack of funding and original carrier degradation. Therefore, solutions to this issue are technically challenging and of considerable cultural and historical importance.

The present research concerns digitization errors in open-reel tapes. The main cause of error is the setting of the tape machine, in particular the choice of the playback speed and equalization standard. This problem is most frequent in cases where a recording contains multiple equalization standards and/or speeds on the same tape. As reported in [18] this issue is prevalent, with 16.7% of open-reel tapes digitized at the Centro di Sonologia Computazionale[1], University of Padova from 2013 to 2020 containing multiple speeds. This paper proposes a correction workflow and digital filters for restoring digitization made with incorrect speeds and equalization standards, therefore providing a tool to save (at least partially) the original content. Following this, perceptions of similarity for these digital filters are assessed through a MUSHRA-inspired test containing 24 participants.

2 SPEED AND EQUALIZATION STANDARDS

Open-reel tapes can be recorded with different speeds: 30 ips ("inches per second", equivalent to 76.2 cm/s), 15 ips (38.1 cm/s), 7.5 ips (19.05 cm/s), 3.75 ips (9.53 cm/s), 1.875 ips (4.76 cm/s) and 0.9375 ips (2.38 cm/s). A tape recorder providing all these speeds in the same machine does not exist [3]. Higher recording/playback speeds are usually adopted by professional machines, such as the one considered in this work: the Studer A810. It covers the four speeds noted above between 30 ips and 3.75 ips.

Another important parameter is the equalization. In analog audio recordings, the equalization curve is used during the recording phase (*pre-emphasis* curve) for extending the dynamic range [9] and improving the Signal to Noise Ratio (SNR) [5] of the recorded signal. During the playback the inverse *post-emphasis* curve is applied in order to restore a flat frequency response.

The magnitude response of the post-emphasis curve (expressed in dB) can be expressed as a combination of two curves with the following formula:

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

where f is the frequency in Hz and t_1, t_2 are the time constants [12]. An alternative mathematical representation of the formula is:

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

where $\omega = 2\pi f$ is the angular frequency in radians [15].

Table 1 shows the time constants adopted in this work. They are the equalization curves used by the Studer A810 and they are the current standards as indicated in [3]. As can be observed, different standards exist for the same speed and this can be a source of error. Additionally, the equalization standard is strictly connected to the speed: usually the curve varies when the speed changes.

In general, an error in the speed setting entails a loss of information and, if not corrected completely, it can compromise the listening experience. Furthermore, an equalization error deeply changes the frequency spectrum of the original signal, compromising its authenticity. Considering the strict relation between speed and equalization, a correct restoration must consider both parameters.

[1] https://csc.dei.unipd.it, last accessed July 23, 2021

Table 1: Equalization filters time constants adopted by the Studer A810.

Equalization	Speed [ips]	t_1 or t_3 [μs]	t_2 or t_4 [μs]
AES (IEC2)	30	∞	17.5
CCIR (IEC1)	15	∞	35
	7.5	∞	70
NAB (IEC2)	15	3180	50
	7.5	3180	50
	3.75	3180	90

3 CORRECTION WORKFLOW

In general, the compensation in the digital domain of speed and equalization errors made during the digitization process of the analog tape should involve the following steps:

(1) The application of the inverse equalization curve used during the reading phase, in order to remove the incorrect curve;
(2) A re-interpretation of the sampling frequency (e.g., changing the original sample rate of a recording from 96 kHz to 48 kHz) in order to obtain the right playback speed;
(3) The application of the correct equalization curve related to the right speed and equalization standard.

Step (2) is not necessary for cases that contain only an equalization error. The re-interpretation of the sample frequency is essential for making the content audible whenever a speed error occurs, but it cannot recover the information that is irrevocably lost during incorrect digitization. Specifically, this loss of information could happen for a digitization performed while reproducing the tape at a speed higher than the one used during the recording phase, since original frequencies are shifted to higher ones that can exceed the audible threshold. The International Association of Sound and Audiovisual Archives (IASA) recommends digitization at a minimum of 96 kHz and 24 bit [3], therefore with this format it is possible to store information up to 48 kHz, the Nyquist frequency. The Studer A810 exceeds the human auditory threshold of 20 kHz and so it is able to read (although not linearly due to hardware limitations) frequency content that would otherwise be lost. In such problematic cases, the information stored in non-audible frequencies is paramount for the restoration of the original content. An alternative to the re-interpretation of the sample frequency could be a sinc interpolation algorithm (not tested in this study).

Figure 1 shows the five steps of the reading and correction process: the first two in the analog domain, the latter three in the digital domain. As indicated in [17, 21], the pre- and post-emphasis curves correspond to the impulse responses of the recording ($w \in L^2(\mathbb{R})$, where $L^2(\mathbb{R})$ is the Lebesgue space of square-summable functions, which is also a Hilbert space) and reproducing ($r \in L^2(\mathbb{R})$) filters, denoted respectively with $\mathcal{W}(x) := x * w$ and $\mathcal{R}(x) := x * r$, where $x \in L^2(\mathbb{R})$ is an analog signal. Therefore, the resulting filter is defined as $\mathcal{E}(x) := \mathcal{R} \circ \mathcal{W}(x) = (x * w) * r$. Considering the

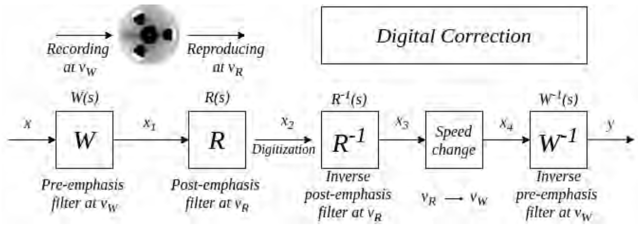


Figure 1: General correction process scheme.

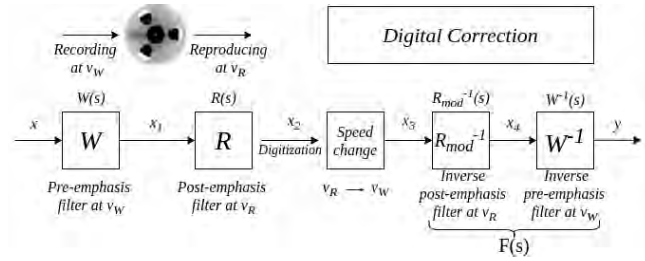


Figure 2: Alternative correction process scheme.

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. $E = id$, 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 case of equalization error, we are dealing with a non-flat equalization $\tilde{\mathcal{E}} = \tilde{\mathcal{R}} \circ \mathcal{W}$, where its transfer function $\tilde{E} = W \cdot \tilde{R} \neq id$, 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: $\tilde{E} \cdot F = id$, where $F : \mathbb{C} \rightarrow \mathbb{C}$ is the transfer function of $\mathcal{F}$. Moreover, the incorrect speed must be considered: to correct a wrong reproducing speed in the digital domain, it is possible to re-interpret the sampling frequency of the signal. This simple process, however, must consider the possibility that filter constants could have changed, due to the definition of the equalization standards presented in Table 1.

Figure 1 also introduces a notation to identify the subsequent manipulations that the signal x undergoes during its elaboration: x_1 refers to the signal recorded on the magnetic tape, therefore it is desired to obtain a signal y which is closest as possible to x by exploiting the information contained in x_1.

To increase the computational efficiency and to easily implement this workflow with technologies such as Web Audio API (where the speed parameter is located in the source node [1]), it is possible to swap the speed change with R^{-1} filter and to design a filter equivalent to the cascade of R^{-1} and W^{-1}, as shown in Figure 2. The design of R^{-1} and W^{-1} filters follows the definition of the standards, which considers a cascade of first order low pass and high pass filters. Therefore, it is possible to show that the transfer function of the reproducing filter is:

$$R(s) = \frac{st_1(1 + st_2)}{1 + st_1} = \frac{st_1 + s^2 t_1 t_2}{1 + st_1} \tag{3}$$

where $s \in C$, while the transfer function of the recording filter is:

$$W(s) = \frac{1 + st_3}{st_3(1 + st_4)} = \frac{1 + st_3}{st_3 + s^2 t_3 t_4} \tag{4}$$

From these definitions, it follows that, when $t_1 = t_3$ and $t_2 = t_4$, $W(s) = R^{-1}(s)$, which corresponds to the case of a correct equalization. In all other cases, it is possible to identify the corrective transfer function as:

$$F(s) = R^{-1} \cdot W^{-1} = \frac{t_3(1 + st_4)(1 + st_1)}{t_1(1 + st_2)(1 + st_3)} \tag{5}$$

Table 2: Combination of speed and equalization standard considered in this study

Case	Recording Speed (ips)	Recording Eq	Reproducing Speed (ips)	Reproducing Eq
A	3.75	NAB	7.5	CCIR
B	3.75	NAB	15	CCIR
C	7.5	NAB	15	CCIR

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

However, this modification must consider the effects of the R^{-1} filter, since in the original schema it operates on just the digitized signal, while in the new one it modifies the re-sampled signal. The result of the two schemes cannot be equal, since in the first case the filter operated on a spectral content altered by the incorrect reproducing speed. Therefore, R^{-1} filter must be substituted by R^{-1}_{mod}, a filter with time constants modified in direct relation with the speed change and considering the definition of the equalization standards presented in Table 1. The general strategy is to multiply the time constants by the reciprocal of the speed change factor which, using the notation introduced in Figure 1, is $m_v = \frac{v_R}{v_W}$.

4 DIGITAL FILTERS

This work aims to create filters for compensating all the different combination of speed and equalization errors during the digitization process. There are 30 possible cases, but in this paper only the three cases presented in Table 2 are considered.

Case A is significant, as the majority of professional or semi-professional tape recorders that are adopted for digitization tasks provide setups with faster speeds, as opposed to 3.75 ips. Regarding Case A, our aim is to test if the proposed correction workflow can compensate the lack of a speed standard in the reproducing tape recorder. Case B is relevant for examples in which larger speed differences (e.g., ×4) occur between the original recorded signal and the digitized one. In this case, considering 96 kHz format, a speed correction through the re-interpretation of sample frequency results in a 24 kHz file, therefore, independently by the tape recorded frequency range, all the frequencies above 12 kHz are lost. For this reason, the proposed method could be useful for speech recordings but not for music. Case C simulates a common eventuality, where there are portions of the same tape recorded in multiple speeds (i.e.

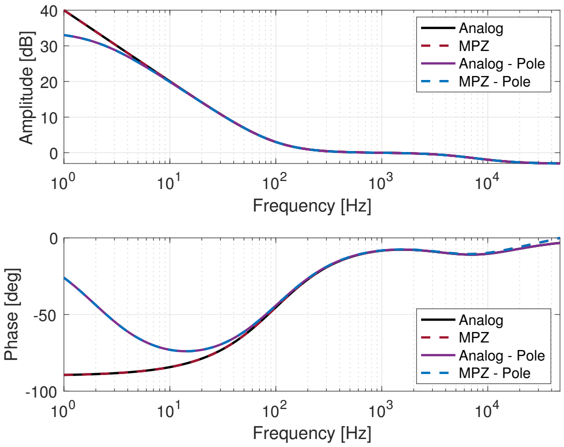


Figure 3: Results obtained with CCIR 30ips recording curve and NAB 15ips reproducing curve. It is possible to notice that the pole translation does not cause evident problems in the digitization of the transfer function.

a tape containing sections recorded at 7.5 and 15 ips, but read at 15 ips) that are not correctly digitized.

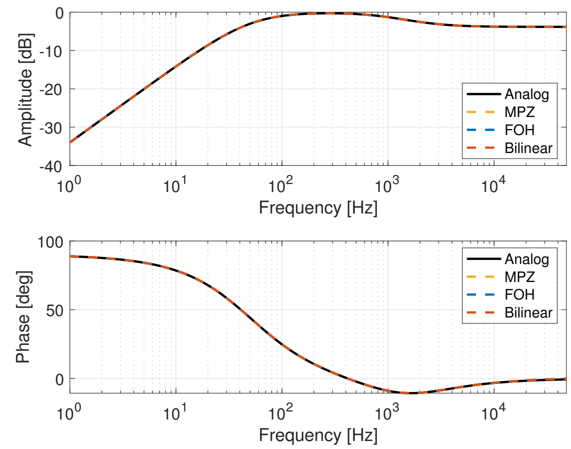
The following description can be applied generally, although in the current study it will only be related to the three filters. The first problem that must be taken into account for the creation of correction filters is their stability: all possible combinations of the four parameters t_1, t_2, t_3 and t_4 must produce stable filters. As can be seen from Table 1, t_1 (and therefore t_3) can assume finite values or can be infinite. As observed in [17], considering Equation 5 as a function with parameters t_1 and t_3, there are four cases:

- $t_1, t_3 < \infty$: no change in the formal structure of (5);
- $t_1, t_3 = \infty$: (5) becomes: $\lim_{t_1, t_3 \to \infty} F(s) = \frac{1+st_4}{1+st_2}$;
- $t_1 = \infty$ and $t_3 < \infty$: (5) becomes: $\lim_{t_1 \to \infty} F(s) = \frac{t_3(1+st_4)}{(1+st_2)(1+st_3)}$;
- $t_1 < \infty$ and $t_3 = \infty$: similarly: $\lim_{t_3 \to \infty} F(s) = \frac{(1+st_4)(1+st_1)}{t_1(1+st_2)}$.

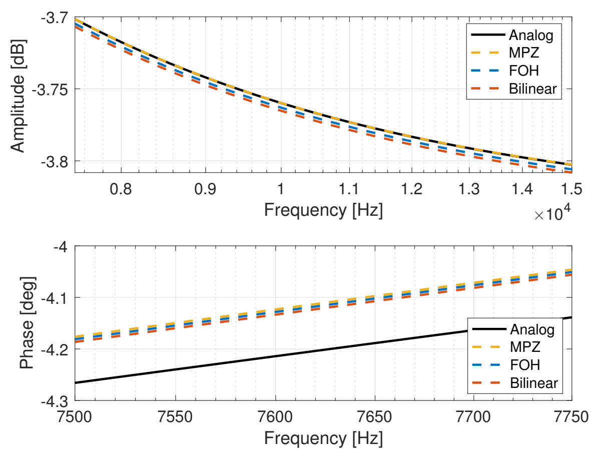
All these filters except the last one are stable as they have poles when $s = -\frac{1}{t_2}$ and/or $s = -\frac{1}{t_3}$, which are both strictly negative. The fourth case gives an unstable filter with a pole in $s = 0$.

The case which corresponds to the unstable filter is relevant in real applications, and so we need to approximate the unstable filter with a stable one which is sufficiently "close" to the first, to produce a similar equalization.

An earlier, related experiment [17] used a simpler design to approach this problem. In the current paper, we instead consider the structure of the transfer function. For cases that are unstable, our approach here was to translate the pole in $s = 0$ to a nearby frequency, so that the overall trend is maintained. A solution was found when the pole was centered at 2 Hz, since it solves the stability problem while altering the audible frequencies only to a small degree. Figure 3 shows the obtained results in one of the possible cases.



(a) Overall trend.



(b) High frequencies particular.

Figure 4: Results obtained with NAB 3.75ips recording curve and CCIR 15ips reproducing curve. In (a), it is possible to notice that all three digitization methods behaves well, since they are all very close to the analog transfer function. However, when zooming in to the high audible frequencies in (b), the MPZ method is the one that best captures the trend of the analog function magnitude response, while it performs worst for phase.

It is possible to notice that, for what concerns the magnitude response, the alterations are all under 20 Hz; however, phase alterations are more visible. It is not completely clear how phase alterations can be perceived [19], since the effects are more or less audible depending on the content of the signal: more for speech, less for music [7]. Future studies could deepen this particular matter.

Now that stability is guaranteed, it is possible to create digital filters using two main approaches [2]: directly designing a digital filter, or starting from the analog domain to design a filter and then transforming or mapping it to the digital domain. In this paper, the second approach was preferred: having the above definitions of the analog filters, with this approach it is possible to easily obtain digital filters having frequency responses similar to the original ones. There are several digitization methods existing in literature. Our decision was made after comparing three of them: the Matching Pole-Zero (MPZ), the Bilinear (or Tustin's method) [10] and the First-Order Hold (FOH)[2]. Figure 4 shows an example, but similar results were obtained for all cases: the MPZ was the best digitization method for what concerns the magnitude response, the Bilinear was the best for phase approximation, while the FOH had performance in the middle among those two. MPZ was chosen, since greater importance was given to the magnitude response. However, subsequent studies will be needed to investigate this particular aspect to verify if this approach is the best one, considering the used samples. Filters were created by using MATLAB® software, after which their impulse response was saved as an audio file in .wav format to be used in a Web Audio API ConvolverNode, which applies a linear convolution effect given an impulse response [1].

5 ASSESSMENT

We conducted an assessment of perception, aimed to evaluate perceivable differences between variants of music and voice excerpts. The design of the experiment was inspired by the MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA) test, a well-established method for evaluating the quality of several variants of an audio stimulus [14, 22]. For our purposes, the MUSHRA-inspired assessment was conducted to quantify differences between a stimulus recorded in magnetic tape and digitized with a correct speed and equalization standard ("Reference") from (a) the same stimulus intentionally digitized with a wrong speed and equalization standard and subsequently fixed by re-interpreting its sampling frequency in order to obtain the correct speed, without applying any other equalization filter ("Foil"), (b) the Reference processed with a low pass filter ("Anchor") and (c) the Foil subsequently corrected with the digital filters proposed in the previous section [17]. Details are provided in Subsection 5.3.

Importantly, while MUSHRA tests typically use a 3.5 kHz low-pass filter as the Anchor (which is at times accompanied by a second Anchor containing a low-pass filter at or close to 7 kHz) [22], here we decided to examine the impact of only a single 7 kHz low-pass filter Anchor. This decision was made based on the findings of prior research [23] which suggests that the use of a 3.5 kHz Anchor is too easy to discern from other variants in a MUSHRA test, and this may lead to a response in which differences between the less-discernible variants become comparatively difficult to perceive [11]. In such a case, we might expect the Anchor to be rated at or near the extreme low end of the rating scale, and ratings for many of the less-discernible variants to occur in close proximity to each other at the opposite end of the rating scale [23]. To combat this, our initial aim was to use a 7 kHz low-pass filter Anchor for all of our stimuli. However, we noted that, due to the comparative lack of low frequencies in spoken voice, for the voice stimuli a 7 kHz Anchor was too difficult to discern from the other variants. Therefore, we used a 3.5 kHz Anchor for voice stimuli and a 7 kHz Anchor for music stimuli. Details are provided in Subsection 5.3.

5.1 Materials

The experiment used 15 audio stimuli: 6 excerpts of popular music, 4 excerpts of electroacoustic compositions, and 5 excerpts of Italian-speech audio. The label "popular" refers broadly to well-known Western styles of music, rather than specifically to Western "pop music". The experiment was presented to participants in three different sections (Set A, Set B, Set C), each with one training stimulus and four assessment samples (see Subsection 5.3). Each excerpt was 10 seconds in duration, and was provided in six different variants, namely:

- "Reference": produced by using the correct equalization standard;
- "Hidden Reference": a copy of the "Reference" but hidden to the participant in the test phase;
- "Anchor": the "Reference" altered with a low-pass filter, with pass band set at 7 kHz for music and 3.5 kHz for speech;
- "Foil": an intentionally incorrect equalization, created by mismatching the recording and reading curves and resampled to the correct speed;
- "Matlab correction": the "Foil" variant corrected by means of a MATLAB® script [17];
- "Web Audio API correction": the "Foil" variant corrected by means of an *ad hoc* web interface adopting Web Audio API, for simulating real-time correction in web application [17].

Both "Reference" and "Foil" variants were recorded and reproduced with a Studer A810. The audio samples of the experiment are available in a Zenodo repository (DOI: 10.5281/zenodo.5121844).

5.2 Participants

Twenty-four participants who were Italian residents (21 male, 3 female) took part in the experiment. Participant age ranged 20-58 years (M = 31.1, SD = 12.9). Participants were asked how many years they had spent playing an instrument or singing (range 5-46 years, M = 17.0, SD = 10.7) and how many years they had spent receiving formal training on an instrument or voice (range 0-20 years, M = 10.2, SD = 5.8).

5.3 Procedures

The experiment was presented to the participants in three different sections (Set A, Set B, and Set C), as outlined below:

(1) Set A contained five music stimuli (Table 3), which were produced by writing a magnetic tape with NAB pre-emphasis curve at 3.75 ips. The Foil variant used an incorrect CCIR post-emphasis curve at 7.5 ips;
(2) Set B contained five spoken-word audio excerpts, with each excerpt being a sentence spoken in Italian coming from the "Orthophonic corpus" of the CLIPS project[3]. The training stimulus was an excerpt spoken by a male, while the test stimuli consisted of two female excerpts and two male excerpts concerning two identical phrases. The samples were

[2]https://it.mathworks.com/help/control/ug/continuous-discrete-conversion-methods.html, last accessed August 27, 2021

[3]clips.unina.it, last accessed August 27, 2021

Table 3: First test groups (Set A). Stimuli with NAB 3.75 ips pre-emphasis curve and CCIR 7.5 ips post-emphasis curve.

Stimulus	Genre	Phase
Richard Wagner *Ride of the Valkyrie*	Popular	Training
Taylor Swift *Shake It Off*	Popular	Test
Queen *We Will Rock You*	Popular	Test
Bruno Maderna *Continuo*	Electroacoustic	Test
Luciano Berio *Différences*	Electroacoustic	Test

Table 4: Third test groups (Set C). Stimuli with NAB 7.5 ips pre-emphasis curve and CCIR 15 ips post-emphasis curve.

Stimulus	Genre	Phase
Carl Orff *Carmina Burana*	Popular	Training
The Weeknd *Save Your Tears*	Popular	Test
Eagles *Hotel California*	Popular	Test
Bruno Maderna *Musica Su Due Dimensioni*	Electroacoustic	Test
Bruno Maderna *Syntaxis*	Electroacoustic	Test

recorded with NAB at 3.75 ips. The Foil variant used an incorrect CCIR post-emphasis curve at 15 ips;

(3) Set C contained five music stimuli (Table 4), which were produced by writing a magnetic tape with NAB equalization at 7.5 ips. The Foil variant used an incorrect CCIR post-emphasis curve at 15 ips.

The web interface for the test was created with BeaqleJS, a framework based on HTML 5 and Javascript [13].

In each set, each stimulus received its own test page (e.g., Figure 5) containing the six variants of that stimulus - Reference, Hidden Reference, Anchor, Foil, Web Audio API correction, and Matlab correction. According to MUSHRA protocol [22], the "Reference" variant was always presented first and labeled, whereas the remaining variants were randomized and unlabeled. The exception to this was the training stimuli, for which all variants were labeled. On each test page, participants were asked to evaluate the Similarity of each presented variant in comparison to the "Reference" variant. Ratings were made via a 100-point rating scale containing: (1-20) corresponding to "Different", (21-40) "Somewhat Different", (41-60) "Slightly Different", (61-80) "Nearly Identical" and (81-100) "Identical". The sets and the stimuli within each set were presented in random orders between participants, to counter any possible ordering effects, although the training stimulus was always presented as the first stimulus in a set.

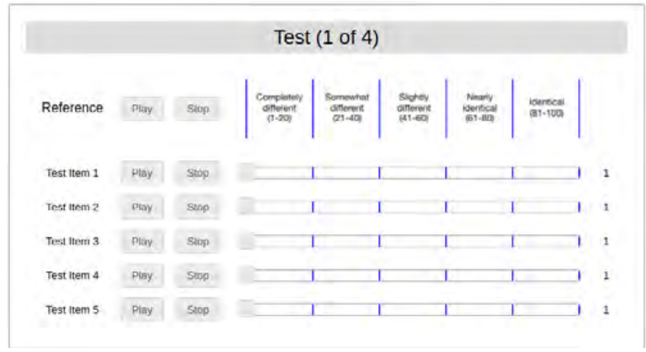


Figure 5: Screenshot of the MUSHRA-inspired test, showing one of the four test samples. The "Reference" is labeled, while "Hidden Reference", "Anchor", "Foil", "Web Audio API correction", and "Matlab correction" are hidden and randomized.

6 RESULTS

While 24 participants took part in the study, some responses were removed prior to analysis after examining the time elapsed on each test page. All cases in which a participant's time on the page was less than 20 seconds were removed, although these were done case-wise rather than removing that participant from the entire dataset. Twenty-three responses were retained for each test page in Set A, 21 responses were retained for each test page in Set B, and 21 responses were retained for each test page in Set C. For each set, a separate within-subjects two-way ANOVA was run, with similarity ratings used as the dependent variable, and containing piece (4 levels) and variant (5 levels, i.e. "Hidden Reference", "Anchor", "Foil", "Matlab correction", "Web Audio API correction") as independent variables. Descriptive statistics for each piece, separated by variant, are reported in Supplementary Table 1 stored in the following Zenodo repository: DOI - 10.5281/zenodo.5118708.

6.1 Set A (Music stimuli)

The Set A ANOVA was significant for both piece ($F(3, 66) = 4.49, p = .006, \eta^2 = .169$) and variant ($F(4, 88) = 71.11, p < .001, \eta^2 = .764$), and produced a significant interaction for piece × variant ($F(12, 264) = 4.18, p < .001, \eta^2 = .160$). Šidák-corrected post hoc tests comparing variants for each piece (see Supplementary Table 2 and Figure 6) indicated that for each piece participants rated the Foil variant significantly lower in similarity than the Hidden reference, and that the 7 kHz Anchor variant was rated significantly lower in similarity for three of four pieces (with the exception being *Continuo*, although this produced a marginally significant result at $p = .055$). Additionally, ratings were not significantly different between the Hidden reference and the Web Audio API correction variant for three of four pieces (with the exception being *Shake it off*), and ratings were not significantly different between the Hidden reference and the Matlab correction variant for all four pieces. This suggests that for Set A both correction methods were effective, although the Matlab variant produced the best result.

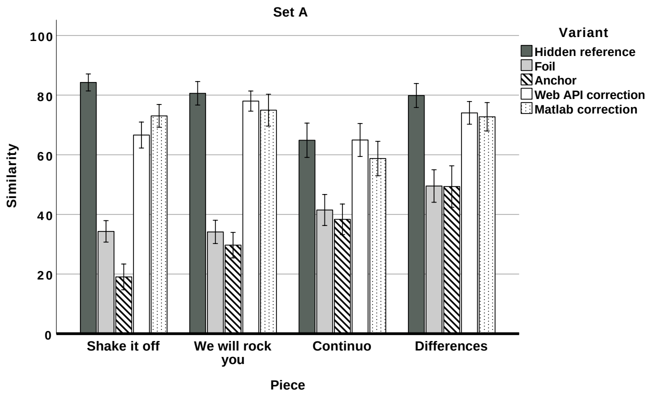


Figure 6: Plotted mean ratings for each stimulus used in Set A, separated by variant. Error bars = +/-1 SE.

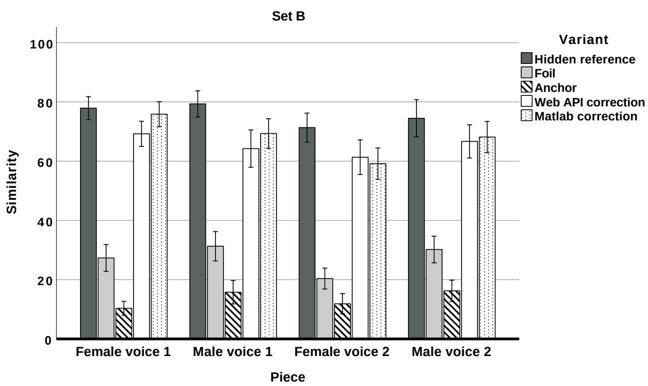


Figure 7: Plotted mean ratings for each stimulus used in Set B, separated by variant. Error bars = +/-1 SE.

6.2 Set B (Voice stimuli)

The Set B ANOVA was significant for both piece ($F(3, 60) = 8.84, p < .001, \eta^2 = .307$) and variant ($F(4, 80) = 83.71, p < .001, \eta^2 = .807$), although the interaction of piece × variant was not significant ($F(12, 240) = 0.91, p = .476, \eta^2 = .044$). Šidák-corrected post hoc tests comparing variants for each piece (see Supplementary Table 2 and Figure 7) indicated that for each piece participants rated both the Foil variant and also the Anchor variant significantly lower in similarity than the Hidden reference. Additionally, ratings were not significantly different between the Hidden reference and either the Web Audio API correction variant or the Matlab correction variant, indicating that both correction methods were effective at compensating for digitization errors for voice stimuli.

6.3 Set C (Music stimuli)

The Set C ANOVA was significant for both piece ($F(3, 60) = 10.98, p < .001, \eta^2 = .354$) and variant ($F(4, 80) = 42.55, p < .001, \eta^2 = .680$), and produced a significant interaction for piece × variant ($F(12, 240) = 8.61, p < .001, \eta^2 = .301$). Šidák-corrected post hoc tests comparing variants for each piece (see Supplementary Table 2 and Figure 8) produced mixed results. These tests indicated that for

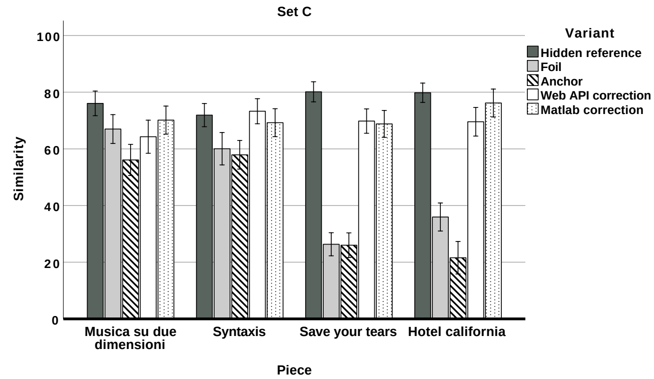


Figure 8: Plotted mean ratings for each stimulus used in Set C, separated by variant. Error bars = +/-1 SE.

the two popular pieces participants rated both the Anchor and Foil variants significantly lower in similarity than the Hidden reference, whereas the two correction variants produced non-significant results, indicating that they were not discernible from the Hidden reference. For the two electroacoustic pieces, none of the variants produced significant differences in similarity compared to the Hidden reference, indicating that participants were not able to reliably distinguish any of the variants from each other for these two pieces. Thus, we cannot infer whether or not the correction variants performed as intended for these two electroacoustic pieces, or not.

7 DISCUSSION AND CONCLUSION

This paper examined a workflow and novel digital filters aimed to compensate errors that occur in the digitization process of open-reel tapes. These errors can occur through a mismatching of the intended equalization standards and playback speeds used in the reading and recording phases, thus impacting the authenticity of the digitized sound and, in some cases, making the content inaudible. The correction workflow and the digital filters aim to produce *ad hoc* compensations for these mismatches, meaning that in cases where it is not possible to re-digitize the original analog audio recordings (which may have deteriorated in the meantime or been lost) they can be used to access the content. In our assessment of perception we examined several variants for a mixture of music and voice stimuli, allowing comparison of the effectiveness of the correction filters for each medium. The stimuli we used also allowed examination of 3 specific mismatches of playback speed and equalization: for Set A, mismatching of music at NAB 3.75 ips and CCIR 7.5 ips; for Set B, mismatching of voice at NAB 3.75 ips and CCIR 15 ips; for Set C, mismatching of music at NAB 7.5 ips and CCIR 15 ips. We also examined the impact of two Anchor variants, with all music stimuli (Sets A and C) containing a 7 kHz Anchor, and the voice stimuli containing a 3.5 kHz Anchor. This inclusion was a necessary because the 7 kHz Anchor was difficult to discern from other voice variants, although prior research has suggested that the use of a 3.5kHz Anchor can lead to a range equalizing bias [23].

Our findings suggest that the Matlab implementation of the correction workflow and digital filters is an effective tool for compensating digitization errors (embodied by the Foil variant), as it was rated statistically identical ($p > .05$) to the Hidden reference variant for all 12 stimuli across all three sets. Similarly, the results suggest that the real-time correction implemented with Web Audio API is an effective tool for compensating these errors, although for one music stimulus (*Shake it off*) this correction variant was rated statistically lower in similarity than the Hidden reference. This suggests that the Matlab correction is slightly more effective than the Web Audio API correction, although further examination is required to tease apart these results. For example, further work could investigate the use of different methods for discretization and frequency warping correction and the implementation of the filters with the Web Audio API *BiquadFilterNode*s [1].

The Foil and Anchor variants were rated significantly lower than the Hidden reference variant for 10 out of 12 stimuli, indicating that the participants were able to reliably differentiate between the incorrectly produced and correctly produced variants more than 80% of the time. However, for the remaining two stimuli, which were the two electroacoustic stimuli used in Set C, the 7 kHz Anchor and the Foil variant were rated as statistically identical to the Hidden reference and the two correction variants. Thus, for these two pieces we cannot make concrete conclusions as to perceptions of the two correction variants. These anomalous results may have been a by-product of the fact that a 7 kHz Anchor was used for the music stimuli, along with the specific stimuli that were chosen. While a 3.5 kHz Anchor may produce a range equalizing biases, it is possible that the use of a 7 kHz Anchor by itself may have led to difficulty in differentiating between variants for certain stimuli. This finding warrants further examination of the impact of various Anchor types in MUSHRA tests, and may also be useful as a cautionary example for future studies that consider the inclusion of only a 7 kHz Anchor. Alternatively, the genre of music used may be able to explain this finding, with excerpts of electroacoustic music seemingly leading to increased difficulty in discerning audible differences between variants. It is possible that many participants were unfamiliar with this style of music, and that unfamiliarity with the compositional textures within the music led to a decrease in perceptual acuity. This interpretation is supported by the results of Set A, in which the popular stimuli produced the clearest statistical differences between the Hidden reference compared with both the Foil and Anchor variants, although further research is required to fully answer this question.

The findings of this study demonstrate the effectiveness of the workflow and digital correction filters across all three proposed cases. In general, when the tape reading speeds were doubled (Sets A and C) it seems that the corrections were perceptually close to the correct digitization, with signals also including high frequencies. In cases of quadruple speed, the results were also close for speech (low and mid frequencies only). In order to confirm these results, additional combinations should be tested in further research. Based on the findings at hand, research in this area can now extend to novel correction filters regarding additional tape speeds, such as 1.875 ips (4.76 cm/s) and 0.9375 ips (2.38 cm/s) speeds, and their related equalizations. Further study could focus on the use of different tape playback devices, as well as evaluating the effects of a slower speed reproduction of a tape recorded with higher speed. The validation of the workflow and digital filters investigated is significant because it can be easily integrated in automatic detection tools [16], restoration platforms [4] and *ad hoc* real-time playback interfaces [6, 8].

REFERENCES

[1] Paul Adenot and Hongchan Choi. 2021. Web Audio API W3C Proposed Recommendation, 6 May 2021 www.w3.org/TR/2021/PR-webaudio-20210506/.

[2] Mohamad Adnan Al-Alaoui. 2007. Novel Approach to Analog-to-Digital Transforms. *IEEE Transactions on Circuits and Systems I: Regular Papers* 54, 2 (2007).

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

[4] Silvia Calamai, Niccolò Pretto, Maria Francesca Stamuli, Duccio Piccardi, Giovanni Candeo, Silvia Bianchi, and Monica Monachini. 2021. Community-Based Survey and Oral Archive Infrastructure in the Archivio Vi.Vo. Project. *Selected papers from the CLARIN Annual Conference 2020* (2021).

[5] Marvin Camras. 1988. *Magnetic Recording Handbook.* Van Nostrand Reinhold Co., New York, NY, USA.

[6] Sergio Canazza, Carlo Fantozzi, and Niccolò Pretto. 2015. Accessing tape music documents on mobile devices. *ACM Trans. on Multimedia Comput., Commun. Appl.* 12, 1s (2015), 20.

[7] Elliot K. Canfield-Dafilou. 2021. *Performing, Recording, and Producing Immersive Music in Virtual Acoustics.* Ph.D. Dissertation. Stanford University.

[8] Carlo Fantozzi, Federica Bressan, Niccolò Pretto, and Sergio Canazza. 2017. Tape music archives: from preservation to access. *International Journal on Digital Libraries* 18, 3 (September 2017), 233–249.

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

[10] François G. Germain and Kurt J. Werner. 2015. Design principles for lumped model discretization using Möbius transforms. In *Proceedings of the 18th International Conference on Digital Audio Effects.* Trondheim, Norway, 371–378.

[11] Harry Helson. 1947. Adaptation-level as frame of reference for prediction of psychophysical data. *The American Journal of Psychology.* 60, 1 (1947), 1–12.

[12] IEC. 1994. 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.

[13] Sebastian Kraft and Udo Zölzer. 2014. BeaqleJS: HTML5 and JavaScript based Framework for the Subjective Evaluation of Audio Quality. In *Linux Audio Conference.* Linux Audio Conference, Karlsruhe, Germany.

[14] David Marston and Andrew Mason. 2005. Cascaded audio coding. *EBU technical review* 304 (2005).

[15] NAB. 1965. Magnetic Tape Recording and Reproducing (Reel-to-Reel).

[16] Niccolò Pretto, Carlo Fantozzi, Edoardo Micheloni, Valentina Burini, and Sergio Canazza. 2019. Computing Methodologies Supporting the Preservation of Electroacoustic Music from Analog Magnetic Tape. *Computer Music Journal* 42, 4 (2019), 59–74.

[17] Niccolò Pretto, Edoardo Micheloni, Anthony Chmiel, Nadir Dalla Pozza, Dario Marinello, Emery Schubert, and Sergio Canazza. 2021. Multimedia Archives: New Digital Filters to Correct Equalization Errors on Digitized Audio Tape. *Advances Multimedia* 1 (April 2021), 15.

[18] Niccolò Pretto, Alessandro Russo, Federica Bressan, Valentina Burini, Antonio Rodà, and Sergio Canazza. 2020. 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.* Torino, Italy, 394–398.

[19] Arkadiy Prodeus, Vitalii Didkovskyi, Maryna Didkovska, and Igor Kotvytskyi. 2017. On Peculiarities of Evaluating the Quality of Speech and Music Signals Subjected to Phase Distortion. In *2017 IEEE 37th International Conference on Electronics and Nanotechnology (ELNANO).* Kiev, Ukraine, 455–460.

[20] Dietrich Schüller. 2001. Preserving the Facts for the Future: Principles and Practices for the Transfer of Analog Audio Documents into the Digital Domain. *Journal of Audio Engineering Society* 49, 7-8 (2001), 618–621.

[21] Roberto Tarantini. 2019. *Realizzazione e valutazione di filtri digitali per l'equalizzazione di nastri magnetici digitalizzati.* Master's thesis. University of Padova.

[22] International Telecommunications Union. 2015. BS. 1534-3. Method for the subjective assessment of intermediate quality level of audio systems. *Recommendation, ITU-R* 10 (2015), 36.

[23] Slawomir Zielinski, Philip Hardisty, Christopher Hummersone, and Francis Rumsey. 2007. Potential Biases in MUSHRA Listening Tests. In *Audio Engineering Society Convention 123.* New York, NY, USA.