A workflow and digital filters for correcting speed and equalisation errors on digitised audio open-reel magnetic tapes

NICCOLÒ PRETTO[1], NADIR DALLA POZZA[1], ALBERTO PADOAN[1], ANTHONY CHMIEL[2],
(niccolo.pretto@dei.unipd.it) (dallapozza@dei.unipd.it) (padoanalbe@dei.unipd.it) (a.chmiel@westernsydney.edu.au)
KURT JAMES WERNER[3], ALESSANDRA MICALIZZI[4], EMERY SCHUBERT[5], ANTONIO RODÀ[1],
(kurt.james.werner@gmail.com) (a.micalizzi@sae.edu) (e.schubert@unsw.edu.au) (antonio.roda@dei.unipd.it)
SIMONE MILANI[1] AND SERGIO CANAZZA[1]
(simone.milani@dei.unipd.it) (sergio.canazza@dei.unipd.it)

[1]*Department of Information Engineering, University of Padova, Padova, IT*
[2]*The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, AU*
[3]*iZotope Inc., Cambridge, Massachusetts, US*
[4]*SAE Institute, Milan, IT*
[5]*Empirical Musicology Laboratory, University of New South Wales, Sydney, AU*

0 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 loss of information (whether partial or complete) due to the degradation of the original signal [1]. 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 [2]. 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 [3] 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 article extends [4], and proposes a correction workflow and digital filters for restoring digitizations made with incorrect speeds and equalization standards, providing a tool to save (at least partially) the original content. Following this, perceptions of similarity for these digital filters are assessed through numerical analyses and a MUSHRA-inspired test containing 24 participants.

1 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 [5]. 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

[1] `csc.dei.unipd.it`, last accessed July 23, 2021

the recording phase (*pre-emphasis* curve) for extending the dynamic range [6] and improving the Signal to Noise Ratio (SNR) [7] 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:

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

where t_1, t_2 are the time constants and $\omega = 2\pi f$ is the angular frequency in radians, where f is the frequency in Hz [8].

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 [5]. 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.

2 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.

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

Step (2) is not necessary for cases that contain only an equalization error. The re-interpretation of the sampling 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 [5], therefore with this format it is possible to store information up to 48 kHz, the corresponding 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. In [4] the authors presented a mathematical notation to describe the process.

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 [9]), 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

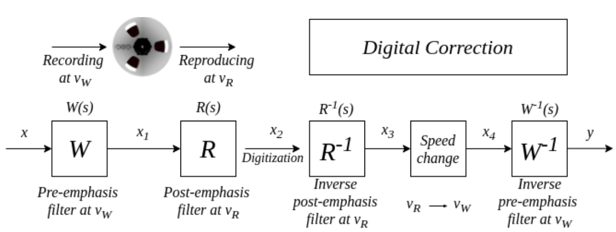


Fig. 1: General correction process scheme.

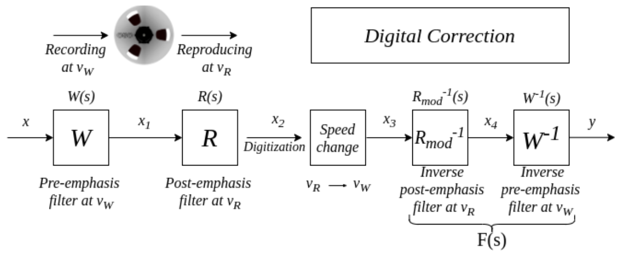


Fig. 2: Alternative correction process scheme.

pass filters. Therefore, it is possible to identify the corrective transfer function as:

$$F(s) = R^{-1} \cdot W^{-1} = \frac{t_3(1+st_4)(1+st_1)}{t_1(1+st_2)(1+st_3)} \tag{2}$$

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 speed change factor which, using the notation introduced in Figure 1, is $m_v = \frac{v_R}{v_W}$.

3 Digital Filters

This work aims to create filters for compensating all the different combinations of speed and equalization errors during the digitization process. Considering the equalization standards definitions in Table 1, it is possible to identify 30 different cases suitable for the application of a correction filter.

When creating such filters, the first problem that must be taken into account 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 [10], considering Equation 2 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 (2);
- $t_1, t_3 = \infty$: (2) becomes: $\lim_{t_1, t_3 \to \infty} F(s) = \frac{1+st_4}{1+st_2}$;
- $t_1 = \infty$ and $t_3 < \infty$: (2) 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 [10] used a simpler design to approach this problem. In the current paper, we instead consider the structure of the transfer function. Our approach here is 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.

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 [11], since the effects are more or less audible depending on the content of the signal: more for speech, less for music [12]. Future studies could deepen this particular matter.

Now that stability is guaranteed, it is possible to create digital filters using two main approaches [13]: 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) [14] and the First-Order Hold

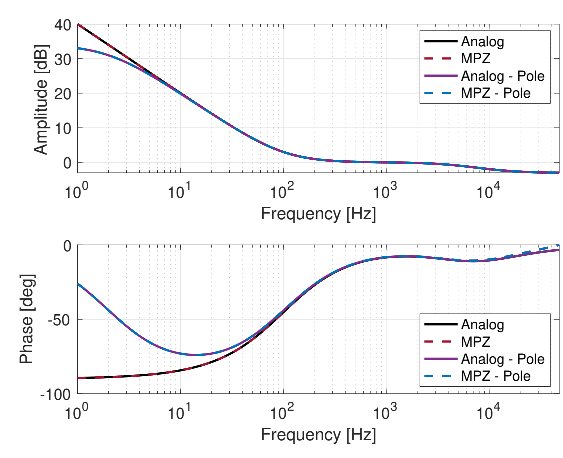
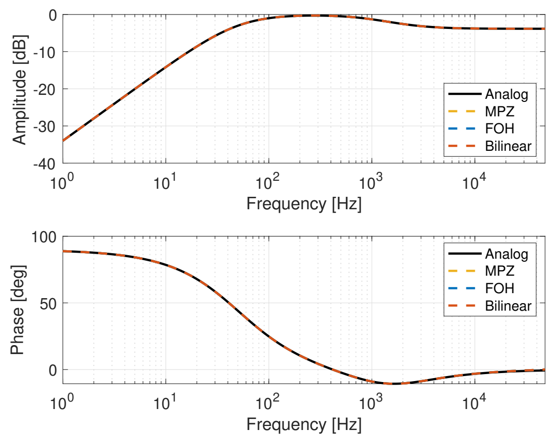


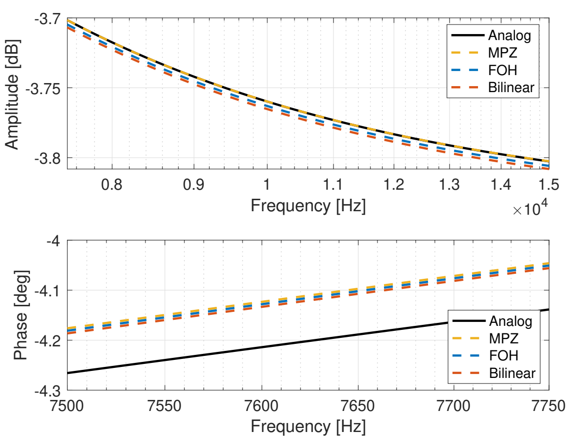
Fig. 3: Results obtained with CCIR 30ips recording curve and NAB 15ips reproducing curve. It is possible to notice that the pole translation does not cause evident problems in the digitization of the transfer function.

(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. For what concerns the following experiment, the MPZ was the chosen method, 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.

[2] `it.mathworks.com/help/control/ug/continuous-discrete-conversion-methods.html`, last accessed 12/11/2021



(a) Overall trend.



(b) High frequencies particular.

Fig. 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.

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 [9].

3.1 Power Spectral Densities

To verify the performance of the filters, we computed the Power Spectral Densities (PSDs) related to the stimuli that will be used in our following Assessment of Perception, by using MATLAB® method `pwelch` with an Hamming window of $N = 1024$ samples with $N/4$ overlapping samples.

An example of the findings is presented in Figure 5, where it is possible to have an idea of the benefits given by the application of the correction filters: the PSD of the corrected variants (both in MATLAB® and Web Audio API applications) are noticeably closer to the Reference variant, when compared to the Incorrect variant. If the paper will be accepted, we will provide the plots for the other cases in a Zenodo repository. With this evidence, we are now ready to set our Assessment of Perception to subjectively verify if the correction is effective.

4 Method for Assessment of Perception

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 [15, 16]. 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 equal-

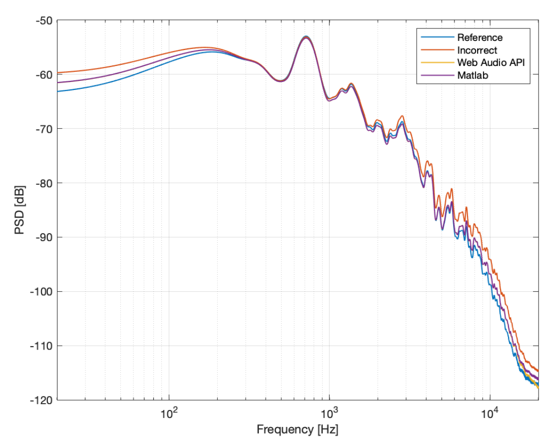


Fig. 5: PSDs of the versions of Carl Orff’s Carmina Burana sample used in our experiment (See Section 4.3).

ization 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 [10]. Details are provided in Subsection 4.3.

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) [15], 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 [17] 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 [18]. 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 [17]. 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 4.3.

4.1 Materials

As it is impractical to examine all 30 cases in a single experiment, is impractical to be tested on a single experiment, therefore we decided to concentrate our study on just three of them, choosing those with most importance. We are going to denote with Case A, B and C, respectively, cases 14, 13 and 9. 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., $\times 4$) 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. a tape containing sections recorded at 7.5 and 15 ips, but read at 15 ips) that are not correctly digitized.

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, B and C, corresponding to the three Cases above), each with one training stimulus and four assessment samples (see Subsection 4.3). 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 [10];
- "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 [10].

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`).

4.2 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 (henceforth "Years playing" - 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 (henceforth "Years training" - range 0-20 years, M = 10.2, SD = 5.8).

4.3 Procedures

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

1. Set A contained five music stimuli (Table 2), 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 recorded with NAB at 3.75 ips. The Foil variant used an incorrect CCIR post-emphasis curve at 15 ips;

[3] `clips.unina.it`, last accessed 12/11/2021

3. Set C contained five music stimuli (Table 3), 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 [19].

In each set, every stimulus received its own test page that was split into two sections. The upper section of the page contained the six variants of that stimulus - Reference, Hidden Reference, Anchor, Foil, Web Audio API correction, and Matlab correction. According to MUSHRA protocol [15], 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. 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. For this upper section of the page, participants were instructed to listen to the Reference variant and the remaining variants in any order and as many times as they wished. The aim was to compare differences in the overall sound between the Reference and each variant, and to rate the Similarity of each variant to the Reference using the provided 100-point rating scale (see Figure 6). Participants were informed that if they had trouble hearing differences between the variants, to focus on the highest and lowest frequencies as this was where the changes should be most apparent.

In the lower section of each page participants rated an additional four variables for the two music sets (Sets A and C), but only an additional one variable for the voice set (Set B). For the music sets, participants rated the Familiarity, Complexity, and Unusualness of the Reference variant, as well as the overall Task difficulty for that entire page. For the voice set participants only rated the overall Task difficulty for that page. Variables such as Familiarity, Complexity, and Unusualness are commonly included in experiments on responses to music stimuli (e.g., [20], [21], [22] and [23]), and so they were included here to help explain anomalous results, and also to allow investigation of whether or not these intrinsic aspects of the music influenced ratings of Similarity. However, as these three variables are not relevant to speech stimuli, they were excluded from Set B. As with the Similarity ratings, these additional responses were each made on a 100-point rating scale as shown in Figure 7. The time that participants spent on each test page was automatically calculated in seconds, and included in the dataset for analysis.

Table 2: 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 3: 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

5 Results and Discussion

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 partic-

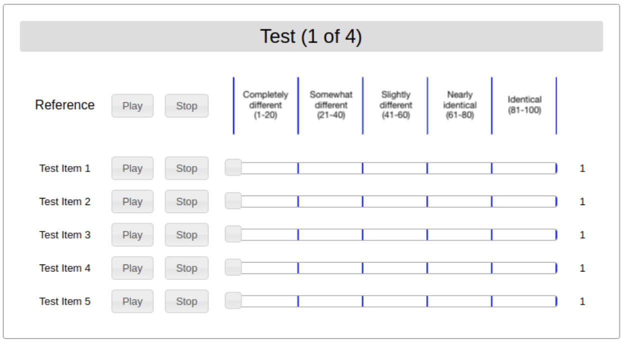


Fig. 6: Screenshot of the MUSHRA-inspired test, showing one of the four test samples. The Reference is labeled, while Hidden Reference, Anchor, Foil, Web Audio API correction, and Matlab correction are hidden and randomized.

ipant'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, twenty-one responses were retained for each test page in Set B, and twenty-one responses were retained for each test page in Set C.

5.1 Analysis of Similarity Ratings by Set, Piece, and Variant

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`. 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 8) 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.

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 9) 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.

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 10) produced mixed results. These tests indicated that for the two popular pieces participants

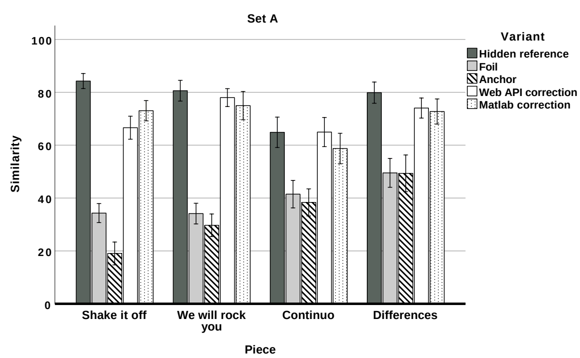


Fig. 8: Plotted mean ratings for each stimulus used in Set A, separated by variant. Error bars = +/-1 SE.

Fig. 7: Screenshot of the MUSHRA-inspired test, showing the Familiarity, Complexity, Unusualness and Task Difficulty ratings.

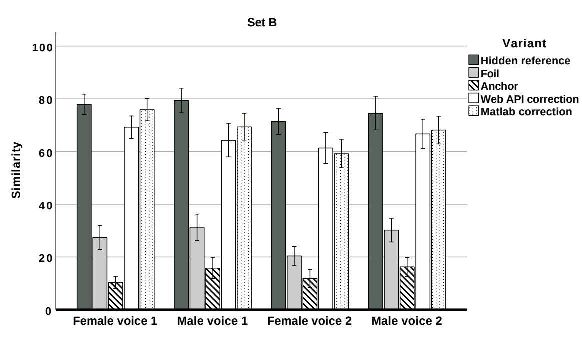


Fig. 9: Plotted mean ratings for each stimulus used in Set B, separated by variant. Error bars = +/-1 SE.

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.

Our findings above 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 and replication is necessary for a thorough comparison. 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 (*Musica su due dimensioni*, and *Syntaxis*), 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

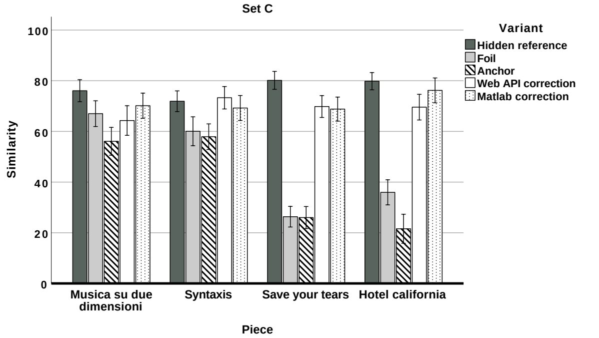


Fig. 10: Plotted mean ratings for each stimulus used in Set C, separated by variant. Error bars = +/-1 SE.

have led to difficulty in differentiating between variants for certain stimuli.

These two anomalous findings in Set C show the need for 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. However, when examining all music stimuli by genre a clear trend is observable in which the four popular stimuli received more "correct" ratings (i.e., the Hidden reference and Correction filters producing higher M values of Similarity, and the Anchor and Foil filters producing lower M values of Similarity) and the four electroacoustic stimuli produced more "incorrect" ratings (i.e., the Anchor and Foil filters producing higher M values of Similarity than they had for the popular stimuli). This trend suggests that the genre of music may play a substantial role in MUSHRA test performance, with electroacoustic music seemingly increasing difficulty to discern audible differences between variants. With this in mind, we next examined ratings of the additional variables between pieces with the aim that these variables may help explain this trend.

5.2 Analysis of Additional Variables by Set and Piece

Descriptive statistics for each additional variable (Familiarity, Complexity, Unusualness, Task difficulty, and Time) are reported in Supplementary Table 3, split by Piece and Set. A MANOVA was performed for each Set; for Sets A and C the dependent variables were Familiarity, Complexity, Unusualness, Task difficulty, and Time, and the independent variable was Piece. For Set B the dependent variables were Task difficulty and Time, and the independent variable was Piece. The results of each MANOVA (consisting of an omnibus test as well as a main effect for each dependent variable) are reported in Supplementary Table 4, and the mean values are also plotted in Supplementary Figure 1. The MANOVAs for Sets A and C each produced a significant omnibus test, as well as significant Main effects for the variables Familiarity, Complexity, Unusualness and Task difficulty; for each of these MANOVAs the variable Time did not produce a significant Main effect. That is, participants spent an equal amount of Time on each test page for Sets A and C. Set B did not produce a significant omnibus test or any significant main effects, and so we conclude that for Set B participants found the task difficulty equal for each piece, and spent an equal amount of Time on each test page.

Šidák-corrected post hoc tests were run for the four significant variables (Familiarity, Complexity, Unusualness and Task difficulty) for Sets A and C. The significance of each test is reported in Supplementary Table 5. Broadly, we can see that the popular music stimuli were rated significantly more familiar, less complex, and less unusual than the electroacoustic music stimuli. However, when two stimuli belonging to the same genre were compared for these variables, 10 out of 12 comparisons were non-significant; only the comparison for Unusualness between *Continuo* and *Differences* and for Familiarity between *Save your tears* and *Hotel california* reached significance. When Task difficulty is examined, all significant comparisons across the two music Sets occurred between stimuli belonging to different genres (i.e., when comparing a popular piece with an electroacoustic piece) and in all of these cases the electroacoustic stimuli were rated significantly higher. With this in mind we can infer that Familiarity, Complexity, and Unusualness of examined music does have a relationship to performance (i.e., rating ability) within an MUSHRA test. Specifically, when we examine ratings increased Familiarity, reduced Complexity, and reduced Unusualness appear to lead to the prevalence of "correct" MUSHRA ratings (as defined above). This suggested relationship is mirrored by the ratings for Task difficulty; stimuli that were less familiar, more complex, and more unusual were rated significantly higher in Task difficulty. Importantly we cannot infer from this analysis the causality of the relationship between these variables and MUSHRA performance; we aim to do this in the following subsection.

5.3 Predictive analysis by variable

In this subsection a series of Multiple Linear Regressions are run, allowing us to examine which variables can significantly predict a "correct" or "incorrect" MUSHRA performance, referring to high ratings of the Hidden reference or Foil variants, respectively. First we perform two analyses with all three sets collapsed: the first analysis used the Hidden reference variant as the dependent variable, and the second analysis used the Foil variant as the dependent variable. For both of these analyses the independent variables were Task difficulty, Time, Age, Years playing, and Years Training. No multicollinearity was detected between these independent variables. Both the analysis on the Hidden reference ($F(8,251) = 5.18$, $p < .001$) with adjusted $R^2 = 0.11$, and on the Foil ($F(5,254) = 15.18$, $p < .001$) with adjusted $R^2 = 0.21$ produced significant ANOVAs.

For each variable the coefficient and significance are reported in Supplementary Table 6. As the analysis on the Foil variant was able to explain a substantially higher proportion of the variance than the analysis on the Hidden reference (as indicated by the R^2 values) we focus on the Foil analysis. Four independent variables (Complexity, Task difficulty, Age, and Years playing) indicated a significant relationship with the Foil variant, whereas Years training was non-significant. Additionally, no significant interactions were observed. As above, Years playing produced the largest coefficient, and as this relationship was negative this indicates that experience in playing a musical instrument helped participants score correctly in the MUSHRA test. Age produced the next largest coefficient, and as this relationship was positive we can infer that older participants performed significantly worse in the MUSHRA test. Additionally, the more difficult a stimulus was perceived, the worse participants scored (i.e., the higher they rated the Foil). Time produced a significant positive relationship although the coefficient was very close to zero, and so this is a much weaker relationship than observed for the other variables.

Next, we performed two similar Multiple Linear Regressions to those above, but limited the data to the two music sets. This enabled us to include additional independent variables that were only collected for the music stimuli, with the complete list of independent variables as Familiarity Complexity, Unusualness, Task difficulty, Time, Age, Years playing, and Years Training. Multicollinearity was observed between Familiarity and Unusualness (r = -.808) and also between Complexity and Task difficulty (r = .748), and so separate analyses were performed with one of these pairs of variables replaced by the other. Both the analysis on the Hidden reference ($F(7,168) = 2.10$, $p = .046$) with adjusted $R^2 = 0.03$, and on the Foil ($F(6,168) = 9.70$, $p < .001$) with adjusted $R^2 = 0.23$ produced significant ANOVAs.

For each variable the coefficient and significance are reported in Supplementary Table 7. As the analysis on the Foil variant was again able to explain a substantially higher proportion of the variance than the analysis on the Hidden reference (as indicated by the R^2 values) we focus on the Foil analysis. Four independent variables (Task difficulty, Time, Age, and Years playing) indicated a significant relationship with the Foil variant, whereas for all other variables $p > .05$. Additionally, no significant interactions were observed. As above, Years playing produced the largest coefficient. As this relationship was negative this indicates that experience in playing a musical instrument helped participants score correctly in the MUSHRA test. Also similar to the earlier analysis, Age produced the second largest coefficient, and as this relationship was positive we can infer that older participants performed significantly worse in the MUSHRA test. Additionally, the more complex and also difficult a stimulus was perceived, the worse participants scored (i.e., the higher they rated the Foil).

With the results of the Multiple Linear Regressions in mind, we conclude that the most important aspect for performing "correctly" in our MUSHRA test for music and voice stimuli was having a high level of experience in playing a musical instrument (but not in training on a musical instrument). Thus, future researchers in this area should aim to match participants as best they can for this variable, and should try to recruit participants with musical experience. Similarly, younger participants performed the best. This might be explained by the gradual decrease in our high-frequency hearing response as we age [24], and researchers should keep this in mind when recruiting. As Complexity and Task difficulty also impacted MUSHRA Performance, researchers should also be careful to balance stimuli for the intrinsic attributes of the stimuli.

6 *A posteriori* Spectral Analysis

6.1 LTAS of Reference and Foil Filters

The final analysis type for this paper examined Long Term Average Spectrum (LTAS) plots, that were produced for the Reference variant and also for the Foil variant for each music stimulus. This approach allows quantification of the spectral differences between these versions, and may give an insight into why participants were able to reliably differentiate between variants for some stimuli (namely the popular pieces) yet why other stimuli (namely the electroacoustic pieces in Set C) produced anomalous results. Each LTAS was a Welch spectrum produced in MATLAB®, using a 256-point Hann window. The "Findpeaks" function was used to take a reading of the Amplitude of the frequency at each interval of 2.5kHz (or as close to that interval as the Findpeaks function would allow). The frequency spectrum for each of the two variants, for each music stimulus, is presented in Supplementary Figures 2 and 3, for Sets A and C respectively. Within the figures you can see the frequency and dB reading at each interval.

Upon cursory visual inspection the popular stimuli appear to contain substantially higher frequencies within the range of 5 to 10kHz, and this is most visually apparent for Set C. Based on this visual examination we hypothesize that when the Foil variant augmented the equalization of each stimulus, it augmented more frequencies in this 5 to 10kHz range for the popular stimuli (especially in Set C) and this led to participants being able to more easily differentiate between variants for the popular stimuli. Thus, in the following spectral analysis we examine dB values at each 2.5kHz interval in an effort to support our visual hypothesis. Two analysis approaches taken, as detailed below.

In approach 1, we compared the differences in amplitude (dB) between the Reference and Foil variants for each stimulus, measured at frequency intervals of 2.5kHz. Following this, a *M* and *SD* difference value between the variants was produced for each piece, shown in Supplementary Table 8. Spectral difference *M* values for the four popular stimuli ranged from 13.0 to 17.0, whereas values of the electroacoustic stimuli ranged from 8.5 to 12.9. This distinction between the styles of music suggests that the popular stimuli received slightly more spectral augmentation than the electroacoustic stimuli, although the difference between music styles is relatively small.

As the majority of the equalization augmentation appears to occur above 5kHz (based on the earlier visual inspection), in approach 2 we only examined spectral differences above 7.5kHz. These difference values are also shown in Supplementary Table 8. The *M* difference values ranged from 6.97 (for *Differences*) to 20.78 (for *Hotel california*). A noticeable difference is evident between three of the electroacoustic stimuli (*Differences*, *Musica su due dimensioni*, and *Syntaxis*) and three of the four popular stimuli, which produced a spectral difference *M* value close to double that of three of the four electroacoustic stimuli. Thus, this spectral analysis supports our visual hypothesis that the popular stimuli received more frequency augmentation from the Foil variant, which is a viable explanation for the anomalous results observed in Set C. Based on this, we recommend that future studies match their stimuli on a spectral level prior to MUSHRA testing in an effort to match stimuli as closely as possible and prevent anomalous results.

6.2 Web Audio API filtering inspection

The results given by the experiment highlighted that there were perceptual differences between the Matlab and Web Audio API correction variants for the stimulus *Shake it off*. We therefore adopted a different Web Audio API correction process to verify if we can improve its performance. Instead of using a ConvolverNode with the correction filter impulse response, we implemented an IIRFilterNode by using the filter transfer function coefficients obtained with the MPZ digitisation method. To compare their performance, we computed the Power Spectral Densities (PSD) estimates of each approach, the Reference and the Foil versions of each stimuli by using MATLAB® method `pwelch` with a Hamming window of $N = 1024$ samples and $N/4$ overlapping samples. In Figure 11 it can be seen that, for example for Bruno Maderna's Continuo stimulus, the performance of the IIRFilterNode is not equal to the one of the ConvolverNode, and it is not clear which Web Audio API correction process performs the best. Therefore, we computed the Root Mean Squared Error (RMSE) between the Reference and (1) the ConvolverNode correction and (2) the IIRFilterNode correction PSD magnitudes for each sample, and then we computed the mean of all RMSEs. Mean PSD RMSE values for the ConvolverNode and IIRFilterNode methods were 2.38 dB/Hz and 2.23 dB/Hz, respectively. When examining the spectral plot for each method, at certain frequencies each approached performed marginally better than the other. However, as the overall differences had an average of 0.003 dB, we conclude that from a perceptual standpoint the two approaches can be considered equal.

6.3 Inspection of Bilinear transform

Following the experiment, we further investigated the performance of the discretization methods by using the RMSE, an objective evaluation method which we set to consider also the phase response of the filters. With this occasion, we also considered the frequency warping effect of the Bilinear transform, which causes the frequency re-

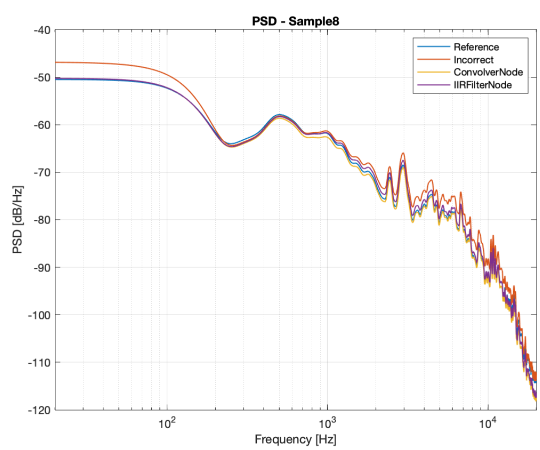


Fig. 11: Inspection of the Web Audio API approaches PSDs.

sponse of the digitized filter to be ”compressed” along the frequency axis. It is possible to compensate this effect by ”pre-warping” frequencies with the following formula:

$$\omega_d = \frac{2}{T}\arctan(\omega_a\frac{T}{2}) \quad (3)$$

where ω_a represents a filter frequency in continuous time, T represents the sampling period and ω_d represents the corresponding filter frequency in discrete time. This compensation is particularly useful when the analog filter presents a salient characteristic, since it permits to match the analog filter frequency response in a specific frequency, but in our analog filters there are none. Nonetheless, we investigated if by prewarping frequencies it is possible to improve the performance of the Bilinear trasform. Therefore, we found the best frequency by choosing the one with the lowest RMSE between the analog filter frequency response and the corresponding bilinear digitization frequency response, obtained by looping the matching frequency in the range of [20, 20k] Hz. Afterwards, we computed the RMSEs between the analog filter frequency response and the MPZ and FOH frequency responses. We found that in all the three cases considered in our experiment the best digitization method is indeed the Bilinear with a prewarping coefficient, presenting RMSEs of 8.6 mW/Hz on average while MPZ RMSEs were of 0.9 mW/Hz on average.

Differently from our evaluation, the RMSE also considers the phase response of the filters, and this means that, overall, the Bilinear with a prewarping coefficient is objectively the best digitisation methods. However, doubts remain when we consider our application field, where the importance of the phase is still uncertain.

6.4 PSD analysis for other cases

As a continuation of the work related to this paper, we decided to examine the performance of the filters in the 27 cases not contemplated by the experiment. We have computed the PSDs of the stimuli by using the same MATLAB® method `pwelch`, but with a wider Hamming window of $N = 4096$ samples and $N/4$ overlapping samples to speed up the computation, since the used samples had a duration of about 6 minutes, if played at the correct speed. Note that these samples contain both music and speech excerpts. Figure 12 shows the PSD of the Reference, Incorrect and Corrected variants of the samples recorded with NAB equalisation at 7.5 ips and reproduced with NAB equalisation at 15 ips. The figure clearly depicts that the two Corrected variants are spectrally closer to the Reference than the Foil is. This indicates that, as intended, both Correction methods are able to alter the Foil variant and produce an outcome that is closer to the correctly produced Reference variant. Based on these embryonic findings, we can expect that, from a perceptual point of view, the action of the filters in these cases could be effective. However, there are also other cases where, in the middle frequencies, the Correct version seems to be more distant from the Reference than the Incorrect, some of them with $m_v = 4$: while this could be related to the loss of information caused by such a great speed difference, this evidence should be nevertheless deeply investigated in further studies also for cases with lower m_vs.

If the paper will be accepted, in the Zenodo repository will be stored the PSDs related to all the considered cases.

7 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 data indicate that participants were not able to differentiate between the Hidden reference variant and Matlab correction variant for all 12 stimuli. The Web Audio API correction variant performed similarly, and could only be differentiated from the Hidden reference variant for one stimulus. While both correction filters provided promising results, additional study with greater sample sizes are needed before more concrete conclusions can be made.

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. The findings of this study demonstrate the effectiveness of the workflow and digital correc-

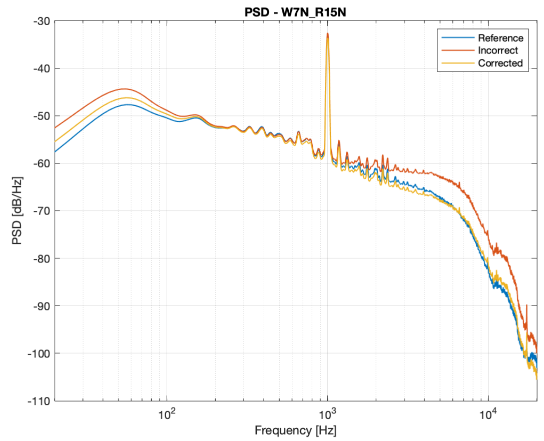


Fig. 12: PSD of Reference, Incorrect and Corrected versions of the samples recorded with NAB equalisation at 7.5 ips and reproduced with NAB equalisation at 15 ips.

tion 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.

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 (Set B) containing a 3.5 kHz Anchor. The use of a 7 kHz Anchor was based on suggestions that a 3.5 kHz Anchor can lead to a range equalizing bias [17], although it was necessary to retain the 3.5 kHz Anchor for the voice stimuli. This was because it was difficult to discern a 7 kHz Anchor from other voice variants due to the lack of low frequencies within the voice stimuli. The use of the 7 kHz Anchor in the music sets led to mixed results; for the popular stimuli the Anchor variant was consistently rated low in Similarity, yet for the electroacoustic stimuli the Similarity ratings for the Anchor were higher than expected. As the Anchor ratings for the electroacoustic stimuli were also higher than those observed in an earlier, related experiment that used a 3.5 kHz anchor for music stimuli [10] we conclude that our inclusion of a sole 7 kHz Anchor negatively impacted MUSHRA performance, specifically for the electroacoustic stimuli. Based on this, we recommend adherence to existing MUSHRA protocols for Anchor variants, being either a sole 3.5 kHz Anchor or both a 3.5 kHz and 7 kHz Anchor in tandem.

Furthermore, the Multiple Linear Regression analyses indicated that the variables Age and Years playing were able to significantly predict performance in the MUSHRA tests. Specifically, increased Age was associated with incorrect scores in the MUSHRA tests, whereas increased Years playing an instrument was associated with correct scores in the MUSHRA tests. This relationship with Age may be due to reduced higher frequency response for older participants, although additional study should aim to confirm this relationship with a larger sample size. Regardless, researchers utilizing the MUSHRA paradigm in future may find it beneficial to focus on younger participants and to prioritize the number of years spent playing an instrument over the number of years spent learning on an instrument. This may also prove useful when considering the prior experience of participants in fields such as audio engineering and mixing live sound. That is, a similar relationship may exist in which years spent working in these fields are a better predictor of MUSHRA performance than years spent training in these fields.

8 REFERENCES

[1] D. Schüller, "Preserving the Facts for the Future: Principles and Practices for the Transfer of Analog Audio Documents into the Digital Domain," *Journal of Audio Engineering Society*, vol. 49, no. 7-8, pp. 618–621 (2001).

[2] C. Fantozzi, F. Bressan, N. Pretto, S. Canazza, "Tape music archives: from preservation to access," *International Journal on Digital Libraries*, vol. 18, no. 3, pp. 233–249 (2017 September).

[3] N. Pretto, A. Russo, F. Bressan, V. Burini, A. Rodà, S. Canazza, "Active preservation of analogue audio documents: A summary of the last seven years of digitization at CSC," presented at the *Proceedings of the 17th Sound and Music Computing Conference, SMC20*, pp. 394–398 (2020).

[4] N. Pretto, N. Dalla Pozza, A. Padoan, A. Chmiel, K. J. Werner, A. Micalizzi, E. Schubert, A. Rodà, S. Milani, S. Canazza, "A Workflow and Novel Digital Filters for Compensating Speed and Equalization Errors on Digitized Audio Open-Reel Tapes," presented at the *Audio Mostly 2021*, AM '21, p. 224–231 (2021), doi:10.1145/3478384.3478409.

[5] K. Bradley, *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) (2009).

[6] 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).

[7] M. Camras, *Magnetic Recording Handbook* (Van Nostrand Reinhold Co., New York, NY, USA) (1988).

[8] NAB, "Magnetic Tape Recording and Reproducing (Reel-to-Reel)," (1965).

[9] P. Adenot, H. Choi, "Web Audio API W3C Proposed Recommendation, 6 May 2021 `www.w3.org/TR/2021/PR-webaudio-20210506/`," (2021).

[10] N. Pretto, E. Micheloni, A. Chmiel, N. Dalla Pozza, D. Marinello, E. Schubert, S. Canazza, "Multimedia Archives: New Digital Filters to Correct Equalization Errors on Digitized Audio Tape," *Advances Multimedia*, vol. 1, p. 15 (2021 April).

[11] A. Prodeus, V. Didkovskyi, M. Didkovska, I. Kotvytskyi, "On Peculiarities of Evaluating the Quality of Speech and Music Signals Subjected to Phase Distortion," presented at the *2017 IEEE 37th International Conference on Electronics and Nanotechnology (ELNANO)*, pp. 455–460 (2017).

[12] E. K. Canfield-Dafilou, *Performing, Recording, and Producing Immersive Music in Virtual Acoustics*, Ph.D. thesis, Stanford University (2021).

[13] M. A. Al-Alaoui, "Novel Approach to Analog-to-Digital Transforms," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 54, no. 2 (2007).

[14] F. G. Germain, K. J. Werner, "Design principles for lumped model discretization using Möbius transforms," presented at the *Proceedings of the 18th International Conference on Digital Audio Effects*, pp. 371–378 (2015).

[15] I. T. Union, "BS. 1534-3. Method for the subjective assessment of intermediate quality level of audio systems," *Recommendation, ITU-R*, vol. 10, p. 36 (2015).

[16] D. Marston, A. Mason, "Cascaded audio coding," *EBU technical review*, vol. 304 (2005).

[17] S. Zielinski, P. Hardisty, C. Hummersone, F. Rumsey, "Potential Biases in MUSHRA Listening Tests," pre-

sented at the *Audio Engineering Society Convention 123* (2007 Oct).

[18] H. Helson, "Adaptation-level as frame of reference for prediction of psychophysical data." *The American Journal of Psychology.*, vol. 60, no. 1, pp. 1–12 (1947).

[19] S. Kraft, U. Zölzer, "BeaqleJS: HTML5 and JavaScript based Framework for the Subjective Evaluation of Audio Quality," presented at the *Linux Audio Conference* (2014 05).

[20] D. E. Berlyne, *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation.* (Hemisphere) (1974).

[21] A. Chmiel, E. Schubert, "Emptying Rooms: When the Inverted-U Model of Preference Fails—An Investigation Using Music With Collative Extremes," *Empirical Studies of the Arts*, vol. 36, p. 027623741773268 (2017 10), doi:10.1177/0276237417732683.

[22] A. Chmiel, E. Schubert, "Unusualness as a predictor of music preference," *Musicae Scientiae*, vol. 23, pp. 426–441 (2019 09), doi:10.1177/1029864917752545.

[23] A. Chmiel, E. Schubert, "Imaginative Enrichment Produces Higher Preference for Unusual Music Than Historical Framing: A Literature Review and Two Empirical Studies," *Frontiers in Psychology*, vol. 11, p. 1920 (2020), doi:10.3389/fpsyg.2020.01920.

[24] L. J. Brant, J. L. Fozard, "Age changes in pure-tone hearing thresholds in a longitudinal study of normal human aging." *The Journal of the Acoustical Society of America*, vol. 88, pp. 813–20 (1990), doi:10.1121/1.399731.

THE AUTHORS

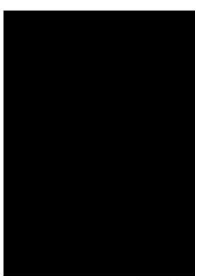

Niccolò Pretto


Nadir Dalla Pozza


Alberto Padoan


Anthony Chmiel


Kurt James Werner


Alessandra Micalizzi


Emery Schubert


Antonio Rodà


Simone Milani


Sergio Canazza

Niccolò Pretto is a Senior Postdoctoral Research Fellow at the University of Padova (IT). His research is primarily focused on sound and music computing, and preservation and access to historical audio documents and cultural heritage in general.

●

Nadir Dalla Pozza is graduated in Computer Engineering at the University of Padova (IT) , where he is a Research Fellow. His interests concern the digitization of cultural heritage, focusing on applications of musicological interest.

●

Alberto Padoan is graduated in ICT for Internet and Multimedia at the University of Padova (IT), where he was also a Research Assistant operating in the digitization of the cultural heritage. He is currently a R&D engineer working in the field of machine learning and optimization problems.

●

Anthony Chmiel is a Postdoctoral Research Fellow at the MARCS Institute for Brain, Behaviour and Development, at Western Sydney University in Australia. His research focuses on aspects of music and psychology such as aesthetics, education, wellbeing, ageing, and computation.

●

Kurt James Werner is a Research Engineer at iZotope who researches virtual analog modeling, artificial reverb, sound synthesis, and the history of music technology. He was formerly an Assistant Prof. of Audio at QUB's SARC and earned his Ph.D. from Stanford Univ.'s CCRMA.

●

Alessandra Micalizzi's biography.

●

Emery Schubert is a Professor in Music at the University of New South Wales Australia and leader of the Empirical Musicology Research Laboratory. He has published over 200 research outputs in music perception, wellbeing, emotion and continuous response.

●

Antonio Rodà is an Associate Professor at the University of Padova (IT). His main research interests are a) affective computing, b) multimodal interactive systems for learning and rehabilitation, c) ICT for musical cultural heritage. He is author of more than 100 papers on national and international journals and peer reviewed conferences.

●

Simone Milani is currently an Assistant Professor with the Department of Information Engineering, University of Padova (Italy). His research interests include digital signal processing, image and video coding, and multimedia forensics. He is author of more than 100 papers on international journals and peer-reviewed conferences.

●

Sergio Canazza is an Associate Professor at the University of Padova (IT). He is director of the Centro di Sonologia Computazionale, and his research interests include affective computing, multimodal interactive systems for learning and rehabilitation, and ICT for musical cultural heritage. He is author of more than 200 papers in international journals and peer reviewed conferences, and CEO of the spin-off Audioinnova srl.