OPEN ACCESS Freely available online

N. Pretto, N. D. Pozza, A. Padoan, A. Chmiel, K. J. Werner, A. Micalizzi, E. Schubert, A. Rodà, S. Milani, and S. Canazza, "A Workflow and Digital Filters for Correcting Speed and Equalization Errors on Digitized Audio Open-Reel Magnetic Tapes" *J. Audio Eng. Soc.*, vol. 70, no. 6, pp. 495–509, (2022 June).
DOI: https://doi.org/10.17743/jaes.2022.0009

A Workflow and Digital Filters for Correcting Speed and Equalization Errors on Digitized Audio Open-Reel Magnetic Tapes

NICCOLÒ PRETTO,[1] *AES Associate Member*, **NADIR DALLA POZZA,**[1] **ALBERTO PADOAN,**[1]
(niccolo.pretto@dei.unipd.it) (dallapozza@dei.unipd.it) (padoanalbe@dei.unipd.it)

ANTHONY CHMIEL,[2] **KURT JAMES WERNER,**[3] *AES Member*, **ALESSANDRA MICALIZZI,**[4]
(a.chmiel@westernsydney.edu.au) (kurt.james.werner@gmail.com) (a.micalizzi@sae.edu)

EMERY SCHUBERT,[5] **ANTONIO RODÀ,**[1] **SIMONE MILANI,**[1] **AND SERGIO CANAZZA**[1]
(e.schubert@unsw.edu.au) (antonio.roda@dei.unipd.it) (simone.milani@dei.unipd.it) (sergio.canazza@dei.unipd.it)

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

This paper presents a workflow and digital filters for compensating speed and equalization errors that can impact digitized audio open-reel tapes. Thirty cases of mismatch between recording and reproducing speed (3.75, 7.5, 15, and 30 in/s) and equalization standards [National Association of Broadcasters (NAB), Consultative Committee for International Radio (CCIR), and Audio Engineering Society] were considered. For three frequent cases of mismatch (NAB 3.75 in/s—CCIR 7.5 in/s; NAB 3.75 in/s—CCIR 15 in/s; and NAB 7.5 in/s—CCIR 15 in/s), MUltiple Stimuli with Hidden Reference and Anchor–inspired tests with $\geq$21 participants assessed the workflow and digital filters, using excerpts of music and voice. Two different correction filters were used, both of which provided promising results. Following this, subsequent analyses examined predictive variables for correct and incorrect MUltiple Stimuli with Hidden Reference and Anchor performance, as well as spectral and numerical differences between filters, which provide key insights and recommendations for further related work.

0 INTRODUCTION

Audio recordings constitute an important part of cultural heritage and priceless source of information for several research areas, such as linguistics, anthropology, and musicology. Nonetheless this heritage risks being permanently lost because of obsolescence, degradation, large numbers, high costs, and short life expectancy [1]. Since analog recordings require physical carriers, 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 [2]. The re-recording issue has been discussed since the 1980s, and several principles are still valid, such as the importance of an accurate and verifiable methodology, the right equipment, and expertise in audio engineering [3]. Nowadays, the digitization process is the only accepted form of re-recording, but the process can introduce artefacts. The active preservation methodology at the base of this work is extensively described in [4, 5], and it differs from Schüller's Type B approach [6] because it does not compensate for unintentional alterations: only intentional alterations such as the equalization curve are compensated during the digitization process.

In recent decades, considerable effort has been made to save large-scale archives, which requires massive digitization projects that often cannot support human supervision

dedicated to ensuring authentic preservation of each audio document. Furthermore many archives report a chronic lack of funding that prevents starting preservation projects with the necessary equipment and expertise. This can lead to digitization errors, which are sometimes not identified until months or years after the task. In these cases, the possible lack of funding and the original carrier degradation could prevent a new digitization project, requiring an urgent solution to this problem. Such solutions are technically challenging and are of considerable cultural and historical importance.

Among all analog carriers, this work concerns digitization errors in open-reel tapes, where the main cause of error is the setting of the tape machine, due to missing information on the original document, in particular the choice of the playback speed and equalization standards. This problem is most frequent in cases where a recording contains multiple equalization standards and/or speeds on the same tape. As reported in [7], 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. In the authors' experience, this is mostly frequent in ethnomusicology and tape music recordings.

This article extends [8] 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 and create access to copies that can be correctly listened to by users. The following sections detail these digitization issues regarding speed and equalization (SEC. 1) and present the correction workflow and digital filters required for this restoration (SECS. 2 and 3, respectively). Following this, in SEC. 4 perceptions of similarity for these digital filters are assessed through statistical analyses and a MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA)–inspired test containing 24 participants. SEC. 5 presents the results obtained by this assessment, and SEC. 6 proposes an *a posteriori* analysis of these findings, which are further discussed in the concluding SEC. 7.

1 SPEED AND EQUALIZATION STANDARDS

Open-reel tapes can be recorded with different speeds: 30 in/s (equivalent to 76.2 cm/s), 15 in/s (38.1 cm/s), 7.5 in/s (19.05 cm/s), 3.75 in/s (9.53 cm/s), 1.875 in/s (4.76 cm/s) and 0.9375 in/s (2.38 cm/s). A tape recorder providing all these speeds in the same machine does not exist [9]. 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 and 3.75 in/s.

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 [10] and improving the signal-to-noise ratio [11] of the recorded signal. During playback, the inverse *post-emphasis* curve is applied in order to restore a flat frequency response.

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

Equalization	Speed [in/s]	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	3,180	50
	7.5	3,180	50
	3.75	3,180	90

The magnitude response of the post-emphasis curve can be expressed (in decibels) as a combination of two curves with the following formula [12]:

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

where t_1 and t_2 are the time constants in seconds and $\omega = 2\pi f$ is the angular frequency in radians, where f is the frequency in hertz. From this equation, the corresponding pre-emphasis curve can be obtained:

$$N(\omega) = 20\log_{10}\left(\frac{1}{\omega t_3}\sqrt{\frac{1+(\omega t_3)^2}{1+(\omega t_4)^2}}\right), \tag{2}$$

where t_3 and t_4 are the time constants in seconds. The notation is different from the post-emphasis curve because, in the following workflow, an equalization error is foreseen, and therefore it is convenient to easily identify the two pairs of time constants.

Table 1 shows the time constants adopted in this work. They are the equalization curves used by the Studer A810, including National Association of Broadcasters (NAB), Consultative Committee for International Radio (CCIR), International Electrotechnical Commission (IEC) and Audio Engineering Society (AES) current standards [9]. 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 the digital domain, the compensation of speed and equalization errors made during the digitization process of the analog tape should involve the following steps:

[1] http://csc.dei.unipd.it/.

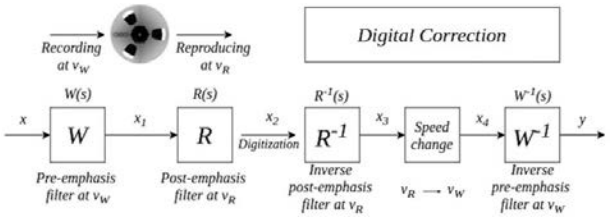


Fig. 1. General correction process scheme.

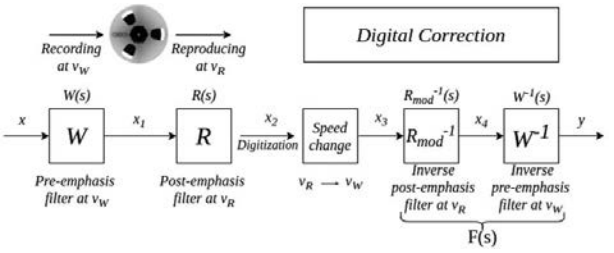


Fig. 2. Alternative correction process scheme.

1. The application of the inverse equalization curve used during the reading phase, 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 to 48 kHz) to obtain the right playback speed; and
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 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 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 recommends digitization at a minimum of 96 kHz and 24 bit [9]; 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, so it is able to read (although not linearly because of 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).

Fig. 1 shows the five steps of the recording, reading, and correction process: the first two in the analog domain and latter three in the digital one. It also introduces a notation to identify the manipulations that the signal x undergoes during its elaboration: x_1 refers to the signal recorded on the magnetic tape, and therefore it is desired to obtain a signal y that is closest as possible to x by exploiting the information contained in x_1. For an extended mathematical notation and description, refer to [8].

The design of R^{-1} and W^{-1} filters follows the definition of the standards, which considers a cascade of first-order low and high-pass filters. To increase the computational efficiency and easily implement this workflow with technologies, such as Web Audio API (where the speed parameter is located in the source node [13]), it is possible to swap the speed change with R^{-1} filter and design a filter equivalent to the cascade of R^{-1} and W^{-1}, as shown in Fig. 2. 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 the R^{-1} filter must be substituted by R_{mod}^{-1}, a filter with time constants modified in direct relation with the speed change and in consideration of the definition of the equalization standards presented in Table 1. The strategy is to multiply the time constants by the speed change factor, which is $m_v = \frac{v_R}{v_W}$; since in Eq. (1) the post-emphasis filter time constants were denoted as t_1 and t_2, the modified time constants $\tilde{t}_1 = t_1 m_v$ and $\tilde{t}_2 = t_2 m_v$. Therefore it is possible to identify the corrective transfer function as

$$F(s) = R_{mod}^{-1} \cdot W^{-1} = \frac{t_3(1+st_4)(1+s\tilde{t}_1)}{\tilde{t}_1(1+s\tilde{t}_2)(1+st_3)}, \tag{3}$$

where $s \in \mathbb{C}$, $\tilde{t}_1$, and $\tilde{t}_2$ are the modified parameters of the reproducing transfer function R and t_3 and t_4 are the parameters of the recording transfer function 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[2] 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 $\tilde{t}_1$, $\tilde{t}_2$, t_3 and t_4 must produce stable filters. As can be seen from Table 1, t_1 (and therefore $\tilde{t}_1$) and t_3 can assume finite values or be infinite. As observed in [14], considering Eq. (3) as a function with parameters $\tilde{t}_1$ and t_3, there are four cases:

- $\tilde{t}_1, t_3 < \infty$: no change in the formal structure of Eq. (3).
- $\tilde{t}_1, t_3 = \infty$: Eq. (3) becomes $\lim_{\tilde{t}_1, t_3 \to \infty} F(s) = \frac{1+st_4}{1+s\tilde{t}_2}$.
- $\tilde{t}_1 = \infty$ and $t_3 < \infty$: Eq. (3) becomes $\lim_{\tilde{t}_1 \to \infty} F(s) = \frac{st_3(1+st_4)}{(1+s\tilde{t}_2)(1+st_3)}$.

[2]The summary of the cases, analysis of the results, images, and impulse responses can be found in the Supplementary Material repository in Zenodo: https://doi.org/10.5281/zenodo.5996876.

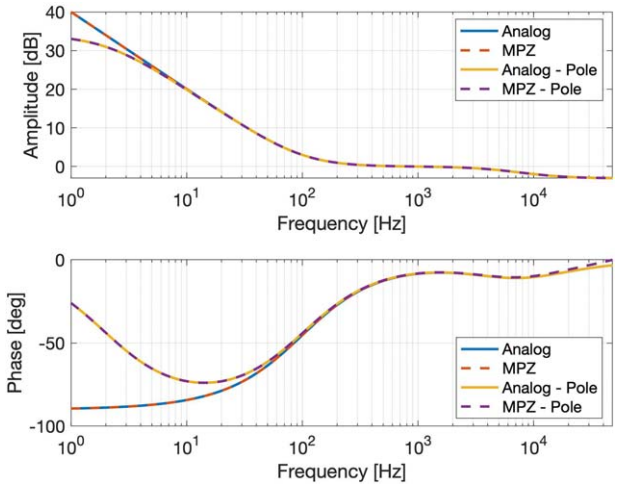


Fig. 3. Results obtained with CCIR 30-in/s recording curve and NAB 15-in/s reproducing curve.

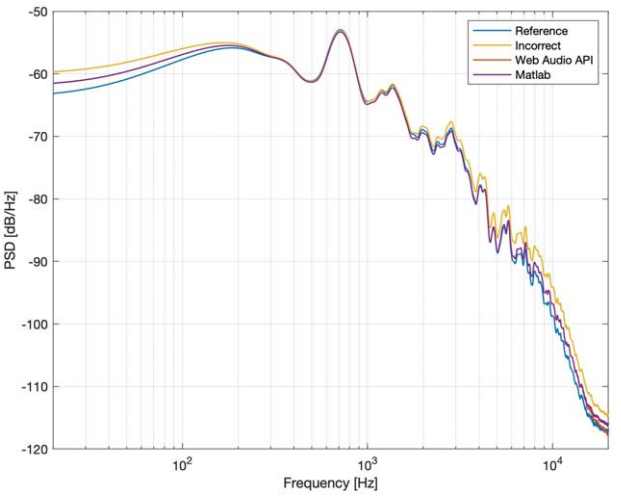


Fig. 4. PSDs of the four variants of Carl Orff's "Carmina Burana" sample used in the experiment (see SEC. 4.3).

- $\tilde{t}_1 < \infty$ and $t_3 = \infty$: similarly $\lim_{t_3 \to \infty} F(s) = \frac{(1+st_4)(1+s\tilde{t}_1)}{s\tilde{t}_1(1+s\tilde{t}_2)}$.

All these filters except the last one are stable because they have poles when $s = -\frac{1}{\tilde{t}_2}$ and/or $s = -\frac{1}{t_3}$, which are both strictly negative. The fourth case gives an unstable filter with a pole in $s = 0$. This last case is relevant in real applications, so the unstable filter needs to be approximated with a stable one that is sufficiently "close" to the first, to produce a similar equalization.

An earlier, related experiment [14] used a simpler design to approach this problem. In the current paper, instead, the structure of the transfer function is considered. The 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. Fig. 3 shows the obtained results in one of the possible cases. When examining Fig. 3, note 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 [15] since the effects are more or less audible depending on the content of the signal: more for speech and less for music [16]. Future studies could examine this aspect in further detail.

Now that stability is guaranteed, it is possible to create digital filters using two main approaches [17]: 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 one was preferred; having the above definitions of the analog filters, with this approach, it is possible to easily obtain digital filters with frequency responses similar to the original ones.

There are several digitization methods that exist in the literature. The decision was made after comparing three of them: the Matching Pole-Zero (MPZ), Bilinear (or Tustin's method) [18], and First-Order Hold (FOH).[3] In general the MPZ was the best digitization method for what concerns the magnitude response, Bilinear was the best for phase approximation, and FOH performance was approximately mid-way between the other 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 verify whether this approach is the best one, considering the used samples.

Filters were created by using MATLAB software,[4] after which their impulse response was saved as an audio file in WAVE format to be used in a Web Audio API ConvolverNode, which applies a linear convolution effect given an impulse response [13]. In this case, since the MATLAB impulse response could become quite long, it was decided to truncate them 10 samples after the quantization error causes the samples to be saved as 0 in 24 bit WAVE format. Since with this approach, some of the impulse responses could become shorter than 0.1 s, it was also decided to have an impulse response at a duration of at least 0.4 s so that information can be stored starting from 2.5 Hz.

3.1 Power Spectral Densities

To verify the performance of the filters, the Power Spectral Densities (PSDs) related to the stimuli that will be used in the following assessment of perception were computed by using MATLAB `pwelch` method with a Hamming window of $N = 1{,}024$ samples with $N/4$ overlapping samples. An example of the findings is presented in Fig. 4, which summarizes the potential benefits given by the application of the correction filters: the PSDs of the Corrected variants

[3] https://it.mathworks.com/help/control/ug/continuous-discrete-conversion-methods.html.

[4] The code of the workflow and the filters are freely available in: https://github.com/CSCPadova/taperecorder_digital_equalizations.

(both in MATLAB and Web Audio API applications) are noticeably closer to the Reference variant, when compared to the Incorrect variant. With this in mind, the assessment of perception is now ready to be set to subjectively verify whether the correction is effective.

4 METHOD FOR ASSESSMENT OF PERCEPTION

An assessment of perception was conducted and was aimed to evaluate perceivable differences between variants of music and voice excerpts. The design of the experiment was inspired by the MUSHRA test, a well-established method for evaluating the quality of several variants of an audio stimulus [19, 20]. For the authors' 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 [14]. Details are provided in SEC. 4.3.

Importantly, although 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) [19], here it was 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 [21] that 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 [22]. In such a case, the Anchor might be expected to be rated at or near the extreme low end of the rating scale, and ratings for many of the less-discernible variants might be expected to occur in close proximity to each other at the opposite end of the rating scale [21].

To combat this, the initial aim was to use a 7-kHz low-pass filter Anchor for all of the stimuli. However it was noted that, because of 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 a 3.5-kHz Anchor was used for voice stimuli, and a 7-kHz Anchor for music stimuli. Details are provided in SEC. 4.3.

4.1 Materials

Because it is impractical to examine all 30 cases in a single experiment, it was decided to concentrate this study on just three of them, choosing those with most importance and denoting them as Case A, B, and C.[5] Case A is significant because 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 in/s. Regarding Case A, the aim is to test whether 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 digitized one. In this case, considering a 96-kHz format, a speed correction through the re-interpretation of sample frequency results in a 24-kHz file; therefore, independently by the tape recorder 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 in/s but read at 15 in/s) that are not correctly digitized.

The experiment used 15 audio stimuli[6]: six excerpts of popular music, four excerpts of electroacoustic music compositions, and five 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 (Sets A, B, and C, corresponding to the three Cases above), each with one training stimulus and four assessment stimuli (see SEC. 4.3). Each excerpt was 10 s long 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 passband 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 re-sampled to the correct speed;
- MATLAB correction: the Foil variant corrected by means of a MATLAB script [14]; and
- 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 [14].

Both Reference and Foil variants were recorded and reproduced with a Studer A810.

4.2 Participants

Twenty-four participants who were Italian residents (21 male and three female) took part in the experiment. Participant age ranged from 20 to 58 years [mean (M) = 31.1, standard deviation (SD) = 12.9]. Participants were asked

[5]Following the numeration of the cases in the Supplementary Materials, they are cases 14, 13, and 9, respectively.

[6]The audio samples of the experiment are available on the following Zenodo repository: https://doi.org/10.5281/zenodo.5996918.

Table 2. Stimuli with NAB 3.75-in/s pre-emphasis curve and CCIR 7.5-in/s post-emphasis curve (Set A).

Stimulus	Genre	Phase
Richard Wagner "Ride of the Valkyries"	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. Stimuli with NAB 7.5-in/s pre-emphasis curve and CCIR 15-in/s post-emphasis curve (Set C).

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

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 in/s. The Foil variant used an incorrect CCIR post-emphasis curve at 7.5 in/s.
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 Corpora e Lessici dell'Italiano Parlato e Scritto (CLIPS) project.[7] The training stimulus was an excerpt spoken by a man, and 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 in/s. The Foil variant used an incorrect CCIR post-emphasis curve at 15 in/s.
3. Set C contained five music stimuli (Table 3), which were produced by writing a magnetic tape with NAB equalization at 7.5 in/s. The Foil variant used an incorrect CCIR post-emphasis curve at 15 in/s.

The web interface for the test was created with BeaqleJS, a framework based on HTML 5 and Javascript [23]. 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 [19], 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.

[7]http://www.clips.unina.it/en/.

The sets and 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 and 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 Fig. 5(a)]. Participants were informed that if they had trouble hearing differences between the variants, they could focus on the highest and lowest frequencies because 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 and the overall Task Difficulty for that entire page. For the voice set, participants rated only 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., [24–27]), and so they were included here to help explain anomalous results and allow investigation of whether or not these intrinsic aspects of the music influenced ratings of Similarity. However, because 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 Fig. 5(b). The time that participants spent on each test page was automatically calculated in seconds and included in the dataset for analysis.

5 RESULTS AND DISCUSSION

Although 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 s 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,

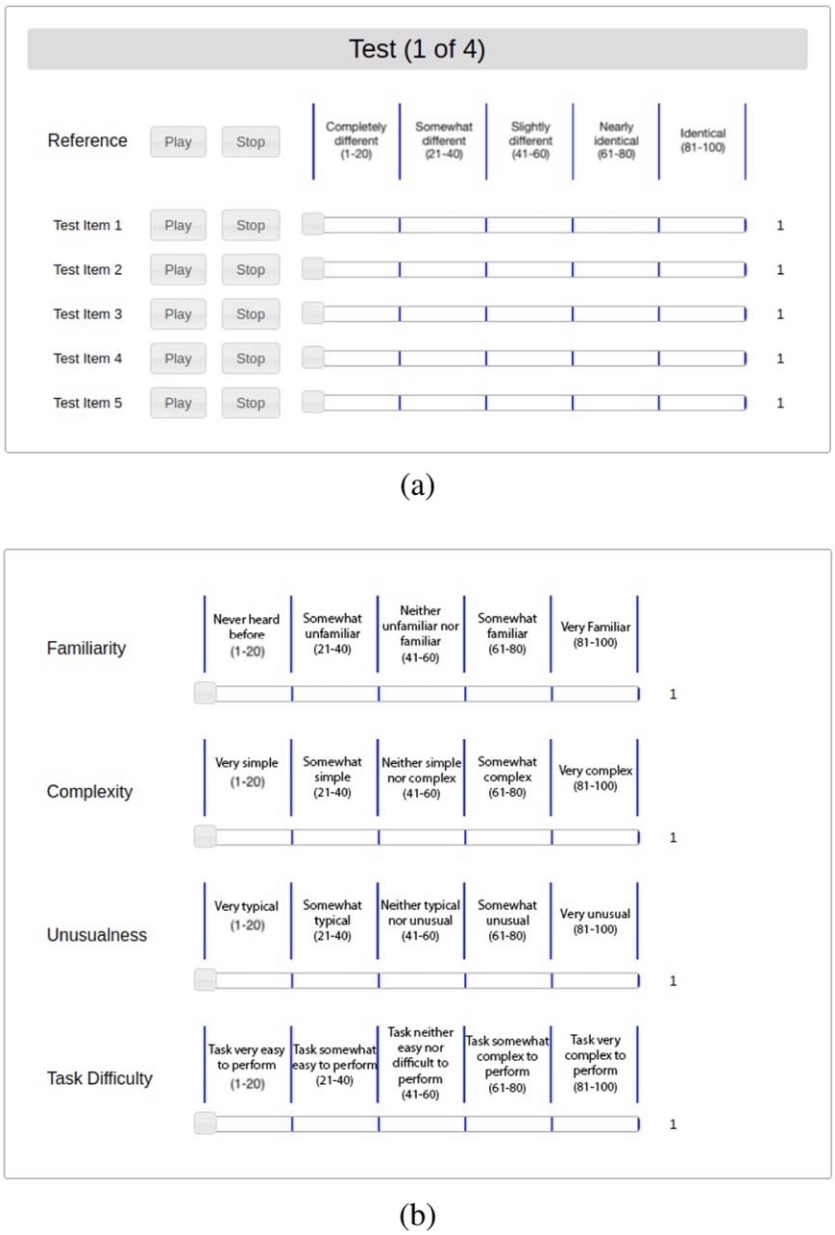


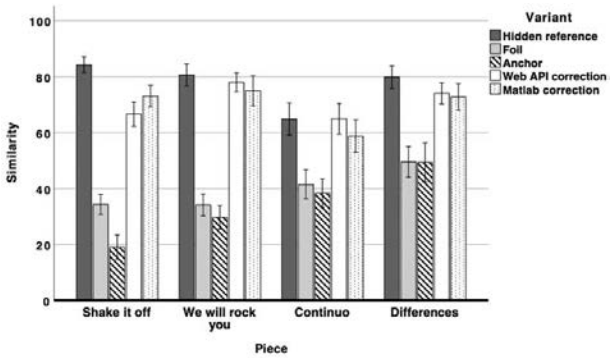
Fig. 5. Screenshot of the MUSHRA-inspired test, showing one of the four test samples. (a) The Reference is labeled, whereas Hidden Reference, Anchor, Foil, Web Audio API correction, and MATLAB correction are hidden and randomized. (b) Familiarity, Complexity, Unusualness, and Task Difficulty rating interface is presented.

21 responses were retained for each test page in Set B, and 21 responses were retained for each test page in Set C.

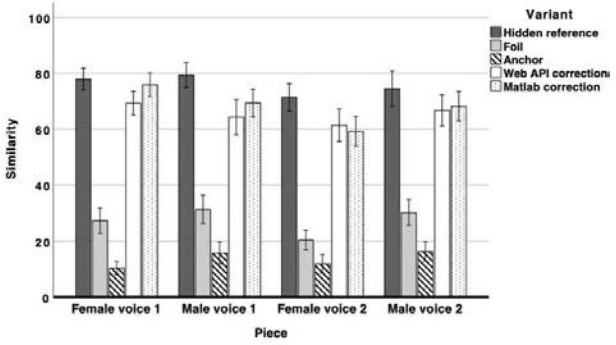
5.1 Analysis of Similarity Ratings by Set, Piece, and Variant

For each set, a separate within-subjects two-way analysis of variance (ANOVA) was run, with Similarity ratings used as the dependent variable and with containing piece (four levels) and variant (five levels, i.e., Hidden Reference, Anchor, Foil, MATLAB correction, and Web Audio API correction) as independent variables. Descriptive statistics for each piece, separated by variant, are reported in Supplementary Table 1 stored in the online repository within the "A Posteriori Analysis" folder.

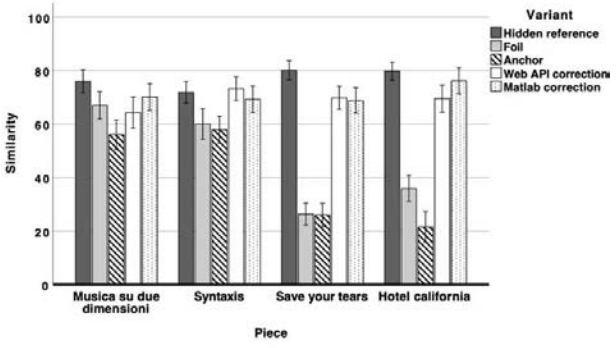
The Set A ANOVA was significant for both piece [$F(3, 66) = 4.49$, $p = 0.006$, $\eta^2 = 0.169$] and variant [$F(4, 88) = 71.11$, $p < 0.001$, $\eta^2 = 0.764$] and produced a significant interaction for piece × variant [$F(12, 264) = 4.18$, $p < 0.001$, $\eta^2 = 0.160$]. Šidák-corrected post hoc tests comparing variants for each piece [see Supplementary Table 2 and Fig. 6(a)] 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 the four pieces (with the exception being "Continuo," although this produced a marginally significant result at $p = 0.055$). Ad-



(a) Set A



(b) Set B



(c) Set C

Fig. 6. Plotted mean ratings for each stimulus used in Set A (a), Set B (b), and Set C (c), separated by variant. Error bars = +/–1 SE.

ditionally ratings were not significantly different between the Hidden Reference and Web Audio API variant for three of four pieces (with the exception being "Shake It Off"), and ratings were not significantly different between the Hidden Reference and MATLAB 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 [$F(3, 60) = 8.84$, $p < 0.001$, $\eta^2 = 0.307$] and variant [$F(4, 80) = 83.71$, $p < 0.001$, $\eta^2 = 0.807$], although the interaction of piece × variant was not significant [$F(12, 240) = 0.91$, $p = 0.476$, $\eta^2 = 0.044$]. Šidák-corrected post hoc tests comparing variants for each piece [see Supplementary Table 2 and Fig. 6(b)] indicated that for each piece, participants rated both the Foil and Anchor variants significantly lower in Similarity than the Hidden Reference. Additionally ratings were not significantly different between the Hidden Reference and either the Web Audio API or MATLAB 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 [$F(3, 60) = 10.98$, $p < 0.001$, $\eta^2 = 0.354$] and variant [$F(4, 80) = 42.55$, $p < 0.001$, $\eta^2 = 0.680$] and produced a significant interaction for piece × variant [$F(12, 240) = 8.61$, $p < 0.001$, $\eta^2 = 0.301$]. Šidák-corrected post hoc tests comparing variants for each piece [see Supplementary Table 2 and Fig. 6(c)] produced mixed results. These tests indicated that for 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 with the Hidden Reference, indicating that participants were not able to reliably distinguish any of the variants from each other for these two pieces. Thus it cannot be inferred whether or not the correction variants performed as intended for these two pieces.

The 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) because it was rated statistically identical ($p > 0.05$) 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 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, concrete conclusions cannot be made 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 chosen specific stimuli. Although a 3.5-kHz Anchor may produce a range equalizing biases, it is possible that the use of a 7-kHz Anchor by itself 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, ratings of the additional variables between pieces were next examined 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 multivariate analysis of variance (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 and main effect for each dependent variable) are reported in Supplementary Table 4, and the mean values are also plotted in Supplementary Fig. 1.

The MANOVAs for Sets A and C each produced a significant omnibus test and 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, so it is concluded 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 it can be seen 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 "Différences" 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, it can be inferred that Familiarity, Complexity, and Unusualness of examined music does have a relationship to performance (i.e., rating ability) within a MUSHRA test. Specifically, when ratings are examined, increased Familiarity, reduced Complexity, and reduced Unusualness appear to lead to the prevalence of "correct" MUSHRA ratings. 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, from this analysis, the causality of the relationship between these variables and MUSHRA performance cannot be inferred; the authors aim to address this in the following section.

5.3 Predictive Analysis by Variable

In this section, a series of Multiple Linear Regressions are run, allowing examination of which variables can significantly predict a "correct" or "incorrect" MUSHRA performance, referring to high ratings of the Hidden Reference or Foil variants, respectively. First, two analyses are performed 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 (in all cases, $r < 0.06$). Both the analysis on the Hidden Reference [$F(8, 251) = 5.18$, $p < 0.001$] with adjusted $R^2 = 0.11$ and on the Foil [$F(5, 254) = 15.18$, $p < 0.001$] with adjusted $R^2 = 0.21$ produced significant ANOVAs.

For each variable, the coefficient and significance are reported in Supplementary Table 6. Because 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), the Foil analysis is focused on. 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 because 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 because this relationship was positive, it can be inferred that older participants performed significantly worse in the MUSHRA test. 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, so this is a much weaker relationship than observed for the other variables.

Next, two similar Multiple Linear Regressions to those above were performed, but the data were limited to the two music sets. This enabled the inclusion of 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 ($r = –0.808$) and also between Complexity and Task Difficulty ($r = 0.748$), so separate analyses were performed with one of these pairs of variables replaced by the other. Both the analysis on the Hidden Reference [$F(7, 168) = 2.10$, $p = 0.046$] with adjusted $R^2 = 0.03$, and on the Foil [$F(6, 168) = 9.70$, $p < 0.001$] with adjusted $R^2 = 0.23$ produced significant ANOVAs.

For each variable, the coefficient and significance are reported in Supplementary Table 7. Because 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), the Foil analysis is focused on. Four independent variables (Task Difficulty, Time, Age, and Years Playing) indicated a significant relationship with the Foil variant, whereas for all other variables, $p > 0.05$. Additionally no significant interactions were observed. As above, Years Playing produced the largest coefficient. Because this relationship was negative, this indicates that experience in playing a musical instrument helped participants score correctly in the MUSHRA test. Similarly to the earlier analysis, Age produced the second largest coefficient, and because this relationship was positive, it can be inferred 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, it is concluded that the most important aspect for performing "correctly" in the 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 try to recruit participants with musical experience. Similarly, younger participants performed the best. This might be explained by gradual decrease in high-frequency hearing sensitivity as people age [28], and researchers should keep this in mind when recruiting. Because 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

In this section, several spectral-based analyses are performed to verify the findings of the experiment and investigate whether the performance of the filters can be improved. The Long Term Average Spectrum (LTAS) has been computed to analyze the differences between Reference and Foil variants; subsequently an inspection of the Web Audio API implementation is presented together with an alternative computational approach. Finally the Bilinear digitization method is deepened with an objective analysis.

6.1 LTAS of Reference and Foil Variants

Here LTAS plots that were produced for the Reference and Foil variants for each music stimulus were examined. This approach allows quantification of the spectral differences between these variants 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 2.5 kHz (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 Figs. 2 and 3, for Sets A and C respectively. Within the figures you can see the frequency and decibel reading at each interval.

Upon cursory visual inspection, the popular stimuli appear to contain substantially higher frequencies within the range of 5 to 10 kHz, and this is most visually apparent for Set C. Based on this visual examination, it is hypothesized that when the Foil variant augmented the equalization of each stimulus, it augmented more frequencies in this 5–10 kHz 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, decibel values at each 2.5-kHz interval are examined in an effort to support the hypothesis. Two analysis approaches were taken, as detailed below.

In approach 1, the differences in amplitude (decibels) are compared between the Reference and Foil variants for each stimulus, measured at frequency intervals of 2.5 kHz. 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.

Because the majority of the equalization augmentation appears to occur above 5 kHz (based on the earlier visual inspection), in approach 2, only spectral differences above 7.5 kHz were examined. These difference values are also shown in Supplementary Table 8. The M difference values ranged from 6.97 (for "Différences") to 20.78 (for "Hotel California"). A noticeable difference is evident between three of the electroacoustic stimuli ("Différences," "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. This spectral analysis supports the visual

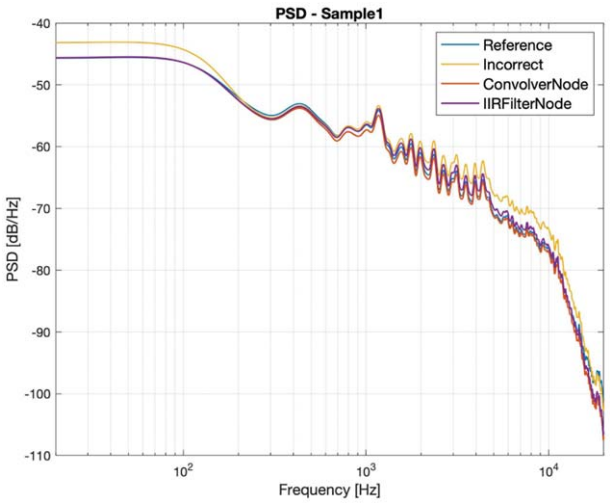


Fig. 7. Inspection of the PSDs Web Audio API approaches.

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, it is recommended that future studies match their stimuli on a spectral level prior to MUSHRA testing in order 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." A different Web Audio API correction process was therefore adopted to verify whether its performance can be improved. Instead of using a ConvolverNode with the correction filter impulse response, an IIRFilterNode was implemented by using the filter transfer function coefficients obtained with the MPZ digitization method. To compare their performance, the PSD estimates of each approach were computed, the Reference and Foil variants of each stimuli by using MATLAB `pwelch` method with a Hamming window of $N = 1{,}024$ samples and $N/4$ overlapping samples.

In Fig. 7 it can be seen that, for the stimulus "Shake It Off," the performance of the IIRFilterNode is not equal to that produced by the ConvolverNode, and it is not clear which Web Audio API correction process performs the best. Therefore the Root Mean Squared Error (RMSE) between the Reference and (1) the ConvolverNode correction and (2) IIRFilterNode correction PSD magnitudes for each sample were computed, and then the mean of all RMSEs were computed. Values for the ConvolverNode and IIRFilterNode methods were 2.38 and 2.23 dB/Hz; based on this, it can be said that the IIRFilterNode method performs generally better. To have a better understanding of this behavior, the difference was compared between (1) the Reference and ConvolverNode PSD and (2) the Reference and IIRFilterNode at each frequency. It was found that at certain frequencies each approach performed marginally better than the other. However, because the overall RMSE difference is only 0.15 dB/Hz, from a perceptual standpoint, the two approaches can likely be considered equal.

When analyzing the plots related to all samples, some considerations can be drawn. In general the ConvolverNode more relevantly modifies the samples when it applies the correction: this could be because of the loss of information caused by the quantization error and consequent truncation of the impulse response of the filters. Moreover, for music samples, sometimes the Incorrect variant shifts toward the Reference variant at high frequencies (above 15 kHz), which is not expected. This could be because of the Studer implementation of the equalization filters. Finally, for voice samples plots, it is possible to see some alterations above 10 kHz: the corrective filters are not capable of correctly restoring the spectral content mainly because of a mix of (1) the non-linearity of the Studer recorder above 40 kHz and (2) the approaching original Nyquist frequency of 48 kHz, which, with $m_v = 4$, now corresponds to 12 kHz.

6.3 Inspection of Bilinear Transform

Following the experiment, the performance of the discretization methods was further investigated by using the RMSE, an objective evaluation method that was set to also consider the phase response of the filters by calculating the complex magnitude after computing the difference between the two filter frequency responses. The frequency warping effect of the Bilinear transform, which causes the frequency response of the digitized filter to be "compressed" along the frequency axis, was also considered. It is possible to compensate this effect by "pre-warping" a filter design [29]. 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 the analog filters there are none.

Nonetheless it was investigated whether, by pre-warping frequencies, it is possible to improve the performance of the Bilinear transform. Therefore the best frequency was found by choosing the one with the lowest RMSE between the analog filter frequency response and corresponding Bilinear digitization frequency response, obtained by looping the matching frequency in the range of 20–20,000 Hz. Afterward the RMSEs between the analog filter frequency response and the MPZ and FOH frequency responses were computed. It was found that in all the three cases considered in this experiment, the best digitization method is indeed the Bilinear with a pre-warping coefficient, presenting RMSEs of 0.86 mW/Hz on average, whereas MPZ RMSEs were of 0.9 mW/Hz on average. Differently from this evaluation, the RMSE also considers the phase response of the filters, and this means that, overall, the Bilinear with a pre-warping coefficient is objectively the best digitization method.

6.4 PSD Analysis for Other Cases

As a continuation of the work related to this paper, the authors decided to examine the performance of the filters in the 17 cases not contemplated by the experiment. The PSDs of the stimuli have been computed by using the same MATLAB `pwelch` method but with a wider Hamming

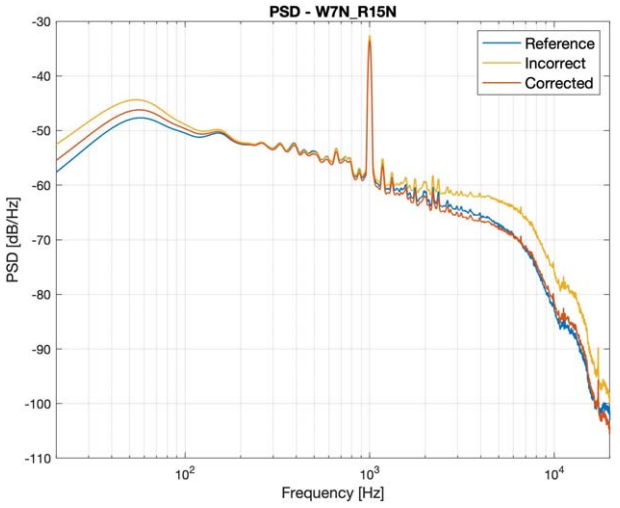


Fig. 8. PSD of Reference, Incorrect, and Corrected variants of the samples recorded with NAB equalization at 7.5 in/s and reproduced with NAB equalization at 15 in/s.

window of $N = 4{,}096$ samples and $N/4$ overlapping samples to speed up the computation, since the used samples had a duration of about 6 min, if played at the correct speed. It was also decided to evaluate the filters only in a MATLAB environment, since testing in a real-time application such as the Web Audio API implementations would be time consuming for such long samples. These examined samples contain both music and speech excerpts.

Fig. 8 shows the PSD of the Reference, Incorrect, and Corrected variants of the samples recorded with NAB equalization at 7.5 in/s and reproduced with NAB equalization at 15 in/s. The figure clearly depicts that the two Corrected variants are spectrally closer to the Reference than the Foil. 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 cursory findings, it can be expected that, from a perceptual standpoint, the action of the filters in these cases could be effective. However there are also other cases where, in the middle frequencies, the Corrected variant seems to be more distant from the Reference than the Incorrect, some of them with $m_v = 4$; although this could be related to the loss of information caused by such a great speed difference, this should be investigated in further studies, and specifically for cases with lower m_vs. The PSDs related to all the considered cases can be found on the online repository.

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 be cased by 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 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 have been lost), they can be used to access the content. Nonetheless this tool is conceived for creating correct access copies [5]; corrected recordings must not replace the preservation copies.

In this assessment of perception, several variants were examined 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. Although both correction filters provided promising results, investigation with greater sample sizes are needed before more concrete conclusions can be made.

The stimuli used also examined three specific mismatches of playback speed and equalization: for Set A, mismatching of music at NAB 3.75 in/s and CCIR 7.5 in/s; for Set B, mismatching of voice at NAB 3.75 in/s and CCIR 15 in/s; and for Set C, mismatching of music at NAB 7.5 in/s and CCIR 15 in/s. 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.

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 [21], 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 because of 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.

Because 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 [14], it is concluded that the inclusion of a sole 7-kHz Anchor negatively impacted MUSHRA performance, specifically for the electroacoustic stimuli. Based on this, adherence to existing MUSHRA protocols for Anchor variants is recommended, 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 decreased performance in the MUSHRA tests, whereas increased Years Playing an instrument was associated with increased performance in the MUSHRA tests. This relationship with Age may be due to reduced higher frequency sensitivity for older participants, although additional study should aim to confirm this relationship with a larger sample size. Regardless, researchers utilizing the MUSHRA paradigm in the future may find it beneficial to focus on younger participants and prioritize the number of years spent playing an instrument over the number of years spent learning 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.

A posteriori spectral analyses were performed to examine anomalous results and extend the conclusions on the performance of each filter. Observed high-frequency augmentation for the Foil variants is of particular note: considering the reduced power in these frequencies for some electroacoustic stimuli, it can explain the anomalous findings by suggesting that participants would have faced difficulty in discerning between variants for these stimuli. Therefore it is imperative that in future studies, stimuli are chosen with care, particularly in matching the spectral content. Following this, an alternative approach for producing a Web Audio API variant was tested to examine whether or not the produced spectral plots varied. Although small differences were observed, it is concluded that these differences would not be perceptible to participants in a MUSHRA test.

In addition to the previously mentioned recommendations for future studies, it is suggested that further work could test the Bilinear with a pre-warp coefficient digitization method, based on the fact it performed the best considering the whole frequency response of the filters. This could also be an occasion to test whether small phase deviations could be relevant to the listening experience. Meanwhile, the performance of MPZ produced filters for cases not verified in this experiment is comparable to the one of the tested filters, and it is therefore viable to proceed with another related assessment of perception. In sum, the provided digital filters are able to provide significant benefit to the ongoing preservation and authentic use of historical audio documents, and subsequent analyses on MUSHRA performance and differences between filters provide key insights and recommendations for further related work.

8 ACKNOWLEDGMENT

A special thanks to the students of the SAE institute of Milan (Italy) and Department of Musicology and Cultural Heritage of the University of Pavia (Italy) that participated in the perceptual experiment described in this paper. Furthermore the authors would like to thank Luca Tacconi of the Sotto il Mare Recording Studios for the recordings of the samples used in this work. Finally the authors would like to thank Dr. Roberto Barumerli, Dario Marinello, Dr. Edoardo Micheloni, Silvio Pol, and Roberto Tarantini, who each contributed to preliminary works. This submission was supported in part by the project *FONTI 4.0* (2105-0020-1463-2019), funded by the Veneto Region, and IT4aREC, funded by the Department of Information Engineering, University of Padova.

9 REFERENCES

[1] M. Casey, "Why Media Preservation Can't Wait: The Gathering Storm," *IASA J.*, vol. 44, pp. 14–22 (2015 Jan.).

[2] D. Schüller, "Preserving the Facts for the Future: Principles and Practices for the Transfer of Analog Audio Documents into the Digital Domain," *J. Audio Eng. Soc.*, vol. 49, no. 7/8, pp. 618–621 (2001 Jul.).

[3] W. Storm, "A Proposal for the Establishment of International Re-Recording Standards," *ARSC J.*, vol. 15, no. 2-3, pp. 26–37 (1983).

[4] F. Bressan and S. Canazza, "A Systemic Approach to the Preservation of Audio Documents: Methodology and Software Tools," *J. Electr. Comput. Eng.*, vol. 2013, paper 489515 (2013 Apr.). https://doi.org/10.1155/2013/489515.

[5] C. Fantozzi, F. Bressan, N. Pretto, and S. Canazza, "Tape Music Archives: From Preservation to Access," *Int. J. Digit. Libr.*, vol. 18, no. 3, pp. 233–249 (2017 Feb.). https://doi.org/10.1007/s00799-017-0208-8.

[6] D. Schüller, "The Ethics of Preservation, Restoration, and Re-Issues of Historical Sound Recordings," *J. Audio Eng. Soc.*, vol. 39, no. 12, pp. 1014–1017 (1991 Dec.).

[7] N. Pretto, A. Russo, F. Bressan, et al., "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*, pp. 394–398 (Torino, Italy) (2020 Jun.). https://doi.org/10.5281/zenodo.3898905.

[8] N. Pretto, N. Dalla Pozza, A. Padoan, et al., "A Workflow and Novel Digital Filters for Compensating Speed and Equalization Errors on Digitized Audio Open-Reel Tapes," in *Proceedings of the Audio Mostly Conference*, pp. 224–231 (Trento, Italy) (2021 Sep.). https://doi.org/10.1145/3478384.3478409.

[9] K. Bradley (Ed.), *Guidelines in the Production and Preservation of Digital Audio Objects: Standards, Recommended Practices, and Strategies* (International Association of Sound and Audio Visual Archives, Johannesburg, South Africa, 2009), 2nd ed.

[10] L. D. Fielder, "Pre-and Postemphasis Techniques as Applied to Audio Recording Systems," *J. Audio Eng. Soc.*, vol. 33, no. 9, pp. 649–658 (1985 Sep.).

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

[12] NAB, "Magnetic Tape Recording and Reproducing (Reel-to-Reel)," *NAB Standard* (1965 April).

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

[14] N. Pretto, E. Micheloni, A. Chmiel, et al., "Multimedia Archives: New Digital Filters to Correct Equalization Errors on Digitized Audio Tapes," *Adv. Multimed.*, vol. 2021, paper 5410218 (2021 Mar.). https://doi.org/10.1155/2021/5410218.

[15] A. Prodeus, V. Didkovskyi, M. Didkovska, and I. Kotvytskyi, "On Peculiarities of Evaluating the Quality of Speech and Music Signals Subjected to Phase Distortion," in *Proceedings of the IEEE 37th International Conference on Electronics and Nanotechnology*, pp. 455–460 (Kyiv, Ukraine) (2017 Apr.). https://doi.org/10.1109/ELNANO.2017.7939796.

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

[17] M. A. Al-Alaoui, "Novel Approach to Analog-to-Digital Transforms," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 54, no. 2, pp. 338–350 (2007 Feb.). https://doi.org/10.1109/TCSI.2006.885982.

[18] F. G. Germain and K. J. Werner, "Design Principles for Lumped Model Discretization Using Möbius Transforms," in *Proceedings of the 18th International Conference on Digital Audio Effects*, pp. 371–378 (Trondheim, Norway) (2015 Nov.).

[19] ITU-R, "Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems," *Recommendation ITU-R BS.1534-3* (2015 Oct.).

[20] D. Marston and A. Mason, "Cascaded Audio Coding," *EBU Tech. Rev.*, no. 304 (2005 Oct.).

[21] S. Zielinski, P. Hardisty, C. Hummersone, and F. Rumsey, "Potential Biases in MUSHRA Listening Tests," presented at the *123rd Convention of the Audio Engineering Society* (2007 Oct.), paper 7179.

[22] H. Helson, "Adaptation-Level as Frame of Reference for Prediction of Psychophysical Data." *Am. J. Psychol.*, vol. 60, no. 1, pp. 1–29 (1947 Jan.). https://doi.org/10.2307/1417326.

[23] S. Kraft and U. Zölzer, "BeaqleJS: HTML5 and JavaScript Based Framework for the Subjective Evaluation of Audio Quality," in *Proceedings of the Linux Audio Conference*, pp. 85–90 (Karlsruhe, Germany) (2014 May).

[24] D. E. Berlyne (Ed.), *Studies in the New Experimental Aesthetics: Steps Toward an Objective Psychology of Aesthetic Appreciation* (Hemisphere Publishing Corporation, Washington, DC, 1974).

[25] A. Chmiel and E. Schubert, "Emptying Rooms: When the Inverted-U Model of Preference Fails—An Investigation Using Music With Collative Extremes," *Empir. Stud. Arts*, vol. 36, no. 2, pp. 199–221 (2018 Jul.). https://doi.org/10.1177/0276237417732683.

[26] A. Chmiel and E. Schubert, "Unusualness as a Predictor of Music Preference," *Music. Sci.*, vol. 23, no. 4, pp. 426–441 (2019 Dec.). https://doi.org/10.1177/1029864917752545.

[27] A. Chmiel and E. Schubert, "Imaginative Enrichment Produces Higher Preference for Unusual Music Than Historical Framing: A Literature Review and Two Empirical Studies," *Front. Psychol.*, vol. 11, paper 1920 (2020 Aug.). https://doi.org/10.3389/fpsyg.2020.01920.

[28] L. J. Brant and J. L. Fozard, "Age Changes in Pure-Tone Hearing Thresholds in a Longitudinal Study of Normal Human Aging," *J. Acoust. Soc. Am.*, vol. 88, no. 2, pp. 813–820 (1990 Aug.). https://doi.org/10.1121/1.399731.

[29] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing* (Pearson, Prentice Hall, Upper Saddle River, NJ, 1999), 2nd ed.

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 (Italy). His research is primarily focused on sound and music computing, preservation and access to historical audio documents, and cultural heritage.

•

Nadir Dalla Pozza graduated with a degree in Computer Engineering at the University of Padova (Italy), where he is a Research Fellow. His interests concern the digitization of cultural heritage, focusing on applications for musicology.

•

Alberto Padoan graduated with a degree in ICT for Internet and Multimedia at the University of Padova (Italy), where he was a Research Assistant. He is currently an R&D engineer working in the field of artificial intelligence.

•

Anthony Chmiel is a Postdoctoral Research Fellow at the MARCS Institute for Brain, Behaviour and Development, at Western Sydney University (Australia). His research focuses on aspects of music and psychology, such as aesthetics, education, wellbeing, aging, 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 Professor of Audio at Queen's University Belfast's Sonic Arts Research Centre and earned his Ph.D. from Stanford University's Center for Computer Research in Music and Acoustics.

•

Alessandra Micalizzi is a Lecturer at the SAE Institute of Milan (Italy). Her research interests are about the practices in participative virtual environments.

•

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 (Italy). He has published over 100 papers in affective computing, multimodal interactive systems for learning and rehabilitation, and information communication technology for cultural heritage.

•

Simone Milani is currently an Associate Professor at the University of Padova (Italy). He has published over 100 papers in digital signal processing, image and video coding, and multimedia forensics.

•

Sergio Canazza is an Associate Professor at the University of Padova (Italy). He is director of the Centro di Sonologia Computazionale and CEO of the spin-off Audio Innova srl. He is an author of more than 200 papers in affective computing, multimodal interactive systems for learning and rehabilitation, and information communication technology for musical cultural heritage.