

A Step toward AI Tools for Quality Control and Musicological Analysis of Digitized Analogue Recordings: Recognition of Audio Tape Equalizations

Edoardo Micheloni, Niccolò Pretto, and Sergio Canazza

Department of Information Engineering (DEI)

University of Padova

`edoardo.micheloni@dei.unipd.it`,

`niccolo.pretto@dei.unipd.it`,

`sergio.canazza@dei.unipd.it`

Abstract. Historical analogue audio documents are indissolubly linked to their physical carriers on which they are recorded. Because of their short life expectancy these documents have to be digitized. During this process, the document may be altered with the result that the digital copy is not reliable from the authenticity point of view. This happens because digitization process is not completely automatized and sometimes it is influenced by human subjective choices. Artificial intelligence can help operators to avoid errors, enhancing reliability and accuracy, and becoming the base for quality control tools. Furthermore, this kind of algorithms could be part of new instruments aiming to ease and to enrich musicological studies.

This work focuses the attention on the equalization recognition problem in the audio tape recording field. The results presented in this paper, highlight that, using machine learning algorithms, is possible to recognize the pre-emphasis equalization used to record an audio tape.

Keywords: audio tape equalization, automatic recognition of physical carrier peculiarities, quality control tool for digitization process, artificial intelligence for musicological analysis

1 Introduction

In the last years, the musicology research field has greatly expanded its original scope by embracing new different research disciplines and methodologies [1]. The potentialities of computer science applied to musicological studies were clear several decades ago when the interdisciplinary domain of computational musicology arose [2], and already in those ages the term artificial intelligence was preponderant. In recent years, research in this field tries to exploit machine learning algorithms in order to obtain meaningful musical concepts and develop models with which to make predictions. Usually these analysis is based on musical features obtained from audio, text or notated score [1].

Unlike born-digital audio files, historical analogue audio documents are indissolubly linked to their physical carriers, on which they are recorded, and to the related audio player (gramophone, tape recorder/player), strongly defining the listening experience [3]. In some case, the peculiarities of the carrier heavily influence musical works and they must be considered during the musicological analysis. However the common analysis, previously described, mainly investigate on musical contents of digital file without considering aspects related to physical carrier.

Nevertheless, scholars can only works on digitized copies of the audio documents because usually original carriers and related playback devices are not available or even missing. Furthermore, these two elements have a short life time expectancy, because of physical degradation and obsolescence, and the only way to maintain the information is to transfer data onto new media and to create a digital preservation copy (active preservation) [4].

Unfortunately, during this process, the history of the document may be distorted and the documentary unit can be broken with the result that the digital copy is not reliable from the authenticity point of view [5]. It usually happens since the process is not completely automatized and sometimes it is influenced by human subjective choices. In this case, artificial intelligence can help operators to avoid errors, enhancing reliability and accuracy of the process. Starting from the analysis of digital copies, AI can discover peculiarities related to the carrier and decide necessary actions to be performed by operators. These kinds of algorithms could be also the base for quality control systems applied to digitization process.

Despite of these problems, the creation of a digital copy can be considered as an opportunity to improve the quality of the musicological analysis. For example, an automatic tool could be useful to investigate the manipulation of the carrier and allow to recreate its history when some information is missing. A step in this direction was done in [5] analyzing video recordings of the tape in order to discover particular elements of the tape itself during the digitization process. On the contrary, in this paper, a study on automatic tools for audio signal analysis is presented, using audio tape recordings as case study. In the Section 2, the peculiarities of this kind of historical audio documents and a first problem to resolve in order to safeguard the authenticity of the preservation copy is described. In the Section 3, the experiment based on most common machine learning techniques is summerised. The results and further developments opened by this work are discussed in Section 4 and 5.

2 Case study: audio tape recordings

Magnetic tape for audio recordings was invented by German Fritz Pfleumer in 1928. Reel-to-reel audio tape recordings rapidly became the main recording format used by professional recording studio until the late 1980s and, for this reason, numerous sound archives preserve large number of audio tapes.

As for every type of analogue carrier, the magnetic tape is also subjected by physical degradation, that can be slowed down but not arrested. So, the digitization process is necessary to prevent that the document reach a level of degradation from which the information is no more accessible [5]. This recording technology is the perfect case study because the constraints imposed by its mechanical and physical limits could be used itself to create music. A clear example is tape music, where the composer becomes also the luthier and the performer of the product recorded on the tape, that can be considered an *unicum* [3]. Furthermore, the magnetic tape is strictly linked to its playback device: the reel-to-reel tape recorder. Before pressing the play button, the machine has to be configured to correctly playback the recordings on the tape and any error implies an audio alteration and the loss of the preservation copy authenticity.

The main two parameters to be configured are reel replay speed and equalization. In this work, only 15 ips (38.1 cm/s) and 7.5 ips (19.05 cm/s) reel replay speed have been considered since they are the most commonly used standards for audio tape.

As far as equalization parameter concerns, during the recording process, the source signal is modified applying an equalization that alter the frequency response (application of a pre-emphasis curve) in order to maximize the SNR of the recorded signal [6]. This alteration has to be compensated during the reading of the tape with a juxtaposition of an inverse curve (post-emphasis curve) in order to obtain the original audio signal. The main standards adopted are CCIR also referred as IEC1 [7], mostly used in Europe and NAB, alternatively called IEC2 [8], mostly adopted in USA. It is important to underline that curves of the same standard can be different according to the reel speed. For example, the cut off frequency of the filter in CCIR differs from 7.5 ips to 15 ips.

Often, speed and equalization standards are not indicated in the carriers. As reported in [9], sometimes any lack of documentation may require the operator to make decisions aurally. The experiment in [10] shows how this task is error-prone. To avoid subjectivity and therefore errors, that can damage the correctness of the preservation copy, the authors proposal is to create a software tool able to discern the correct equalization. This solution is useful not only to aid operators in the digitization process, but can be useful also for musicologists: if they study a digitized copy of unknown provenance they can prove the correctness of the digitization and, if necessary, compensate the error.

3 Equalization recognition

This work wants to prove that machine learning algorithms are able to recognize equalizations using features extracted from small samples of a digitized tape. The experiment is based on four datasets developed in laboratory. They are composed by samples that cover all the combinations of right and wrong chain of filters that can occur while audio tapes are digitized (Tab.1). The samples are characterized by two speeds: 7.5 and 15 ips. For each of the two speeds, white noise has been recorded on half of the samples, while the remaining have been

Table 1. Characterization of the four dataset used in the experiment regarding audio content and recording/reeding speed

Recording/Speed	7.5 ips	15 ips
Silence	dataset A	dataset C
White noise	dataset B	dataset D

recorded with a silence track (silence were recorded on the virgin tape and then acquired). Every dataset contains four type of samples made alternating CCIR and NAB equalization in pre- and post-emphases. The four resulting pairs are CCIR-CCIR (CC), NAB-NAB (NN), CCIR-NAB (CN) and NAB-CCIR (NC). In other words, the first two pairs have the correct juxtaposition of the recording equalization with the writing one, while the other pair is reading the tape with an uncorrected equalization. In this analysis, combination between the two speed has not been taken into account (i.e., NAB at 7.5 ips with CCIR at 15 ips). The samples have been obtained using always the same machine and recorded onto two virgin tapes. Every dataset is composed by 1200 samples with a duration of one second: 300 samples for each categories. With the Matlab tool Mirtoolbox (Music Information Retrieval Toolbox [11]), 13 Mel-Frequency Cepstral Coefficients (MFCCs) has been extracted. These features, originally developed for speech-recognition systems, have given good performance for a variety of audio classifications [12] and they allow a low computational cost and a fast training. For these reasons, the vectors of 13 coefficients have been chosen for the machine learning algorithms. The objective is to evaluate if these algorithms are able to discern automatically the samples and to group them in different clusters/classes.

The experiment is divided in two steps: cluster analysis and classification. The first step exploits the two main methods of cluster analysis (unsupervised learning): hierarchical clustering and K-means clustering. In the first method, different distance measures (i.e. euclidean, chebychev, cosine, etc.) and linkage methods (i.e. average, single, etc.) have been used (with the constraint of maximum four clusters) while, in the second, the parameters were distance measures and number of clusters (cluster from 2 to 4). The number of different combinations for the first method is 188 (47 x 4), whereas for the second is 48 (12 x 4).

The second step exploits three of the most common techniques of supervised learning: Decision Tree, K-Nearest Neighbors, Support Vector Machine (SVM). Concerning the first technique, three presets of classifier have been used and they mainly differ for the maximum number of splits: Simple Tree (maximum 4 splits), Medium Tree (maximum 20 splits) and Complex Tree (maximum 100 splits). The SVM has been used in five variants which differ for the kernel function: Linear, Quadratic, Cubic, Fine and Gaussian. The Nearest Neighbors classifier has been tested in six variants which differ for number of neighbors and distance metric: Fine, Medium, Coarse, Cosine, Cubic and Weighted. K-Fold Cross-Validation (with $k = 4$) is the model validation technique used for the experiment. Every

dataset has been divided in a training set with the 75% of the cepstral coefficients vectors available and a test set with the other 25% of the samples and each group of test are analyzed with the twelve classifiers described above.

4 Results

4.1 Clustering results

The preliminary results are obtained from the clustering analysis and are the following:

- in the case of white noise recordings, dataset B and D, is possible to highlight a first cluster containing the samples generated with the right chain of filters (NN, CC), a second containing one of the wrong juxtaposition of filters (NC) and a third with the other wrong juxtaposition;
- in the case of silence tracks, is possible to identify a cluster describing samples with NAB post-emphases filter and another describing samples with CCIR post-emphases filter.

Most of the different combinations of distances and linkage methods of Hierarchical clustering are able to discern white noise samples. Tab.2 presents an example of good result obtained with Hierarchical clustering. In general, K-means does not work for this kind of samples, excepting for the algorithm that use distance emph cityblock that is able to discern the three clusters. Vice versa, the opposite trend can be observed for silence samples, where K-Means algorithms achieved good results using most of the distances, while Hierarchical is able to divide samples in only few exceptions. An example could be observed in Tab.3. One further observation to point out is that there are few differences between the clustering obtained from 7.5 ips and 15 ips samples. In general, this result was expected, since the only differences are in the cut-off frequency in CCIR equalization (from 2kHz to 4kHz) and this should not compromise the analysis [7].

While the clusterings obtained from the white noise recordings were expected, the one obtained by the silence tracks can be explained with [13], where Mallinson analysis found that the dominant noise source in modern tape recorders is mostly originated from both the reproduce head and the recording medium itself and not from the write head. Therefore, in the case of silence samples, the background noise due to the write head is not powerful enough to be discerned from the one generated from the reading one.

Table 2. Four clusters of white noise samples resulting from Hierarchical clustering with Euclidean distance and centroid as linkage methods

Cluster	Cluster 1				Cluster 2				Cluster 3				Cluster 4			
Distance	CC	CN	NC	NN	CC	CN	NC	NN	CC	CN	NC	NN	CC	CN	NC	NN
# samples	0	0	2	0	0	0	298	0	0	300	0	0	300	0	0	300

Table 3. Two clusters of silence samples resulting from K-Means clustering with squared Euclidean distance

Cluster	Cluster 1				Cluster 2			
distance	CC	CN	NC	NN	CC	CN	NC	NN
# samples	8	300	8	299	292	0	292	1

4.2 Classification results

In this experiment, the K-Fold Cross Validation is used to evaluate the capability of the model to divide the dataset in:

1. correct equalization and wrong equalization;
2. correct equalization, CN, NC;
3. all four pairs of pre- and post-emphases juxtaposition;
4. post-emphases curves.

The last group of test has been added considering the results of the first step. In fact, the objective of this work is to detect the pre-emphases curve but the results obtained in the first step highlight the possibility to detect the post-emphases equalization for the silence track. The results of classification confirm the one of clustering analysis: the noisy datasets allow to detect the correct equalization and discriminate between the two wrong chain of filters, whereas the silence dataset is useful to detect only the post-emphases curve. Even in this case there are no differences between 7.5 ips and 15 ips.

To be more precise, in the first two groups of tests, the indexes of performance of the classification are 1 or very close to that for white noise samples. In both the datasets, the best classification is obtained with the Decision Tree classifiers (simpleTree, mediumTree, complexTree) where the indexes of Accuracy, Recall, Specificity are exactly 1.

In the third group, for 15 ips samples the results show indexes are equal or near to one for CN and NC, but not for CC and NN classes. In other words, the classifiers correctly recognize the wrong equalization pairs but have some difficulties to discern the correct pairs (CC, NN), confirming the results obtained with clustering analysis. For 7.5 ips, an unexpected result arise with cubicSVM on white noise samples dataset: the indexes are 1 for CN and NC classes and tend to the same value for the CC and NN classes. In other words, the classifier is able to recognize all the four type of samples. More details are shown in Tab.4 , where the Accuracy of the classification is 0.97. This result could be explained by non ideal analogue filters or small misalignment in the calibration procedure.

In the last group, the best results are obtained with cubicSVM on the silence samples dataset. As expected from the clustering analysis the silence samples allow to precisely detect the post-emphases equalization. As in the first two group of test the indexes of Accuracy, Recall, Specificity are exactly 1.

Table 4. Indexes of the classification with the four combination of filters on white noise samples using cubic SVM. The accuracy of this test is 0.97

filtersChain	Recall	Specificity	Precision
CC	0.907	0.996	0.986
NC	1	1	1
CN	1	1	1
NN	0.987	0.969	0.926

5 Conclusions and future works

This paper highlights the main problems concerned the physical carriers of analogue audio documents during the digitization process and the musicological analysis. The strictly link between carrier and content defines the listening experience, therefore it is important to preserve it in the digital copy. The creation of a correct preservation copy require firstly the certainty of the correct configuration of the reply machine. This step is not easy to accomplish due to different standards used for tape recorders. In this case, AI tools can simplify the work of operators, helping them in some decisions that must be taken during the digitization process and becoming the base of quality control systems. Furthermore, they could be part of new instruments aiming to ease and to enrich musicological studies.

The results of the preliminary study presented in this paper, highlight that, using machine learning algorithms, is possible to recognize the pre-emphasis equalizations used to record the tapes. This allows to use the correct inverse equalization during the digitization process, balancing the recording equalization and obtaining the original sound.

This encouraging result, obtained from recordings of white noise and silence tracks recorded in laboratory, open the way to further experiments with “real” datasets with samples extracted directly from historical audio recordings. The data collected from this new dataset could be used to compare the results obtained with the ones from [10], to have a comparison between human and artificial classification. In addition, a further work could be the study of additional features to increase the performance of the AI algorithms with more information on the spectral behaviors. This is only a small step toward the development of AI tools for quality control and musicological analysis of digitized analogue recordings, but can surely considered a not negligible first step.

6 Acknowledgments

The authors would like to thank Fabio Casamento, who contributed to the coding of the Matlab algorithms, Valentina Burini and Alessandro Russo, who contributed to the creation of the datasets, and Giorgio Maria Di Nunzio for the several helpful suggestions.

References

1. Serra Xavier. The computational study of a musical culture through its digital traces. *Acta Musicologica*, 89(1):24–44, 2017.
2. Bernard Bel and Bernard Vecchione. Computational musicology. *Computers and the Humanities*, 27(1):1–5, Jan 1993.
3. Sergio Canazza, Carlo Fantozzi, and Niccolò Pretto. Accessing tape music documents on mobile devices. *ACM Trans. Multimedia Comput. Commun. Appl.*, 12(1s):20:1–20:20, October 2015.
4. Federica Bressan and Sergio Canazza. A systemic approach to the preservation of audio documents: Methodology and software tools. *JECE*, 2013:5:5–5:5, January 2013.
5. Carlo Fantozzi, Federica Bressan, Niccolò Pretto, and Sergio Canazza. Tape music archives: from preservation to access. *International Journal on Digital Libraries*, 18(3):233–249, Sep 2017.
6. Marvin Camras. *Magnetic Recording Handbook*. Van Nostrand Reinhold Co., New York, NY, USA, 1987.
7. IEC. Bs en 60094-1:1994 bs 6288-1: 1994 iec 94-1:1981 - magnetic tape sound recording and reproducing systems — part 1: Specification for general conditions and requirements, 1994.
8. NAB. Magnetic tape recording and reproducing (reel-to-reel), 1965.
9. Kevin Bradley. *IASA 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, 2009.
10. Valentina Burini, Federico Altieri, and Sergio Canazza. Rilevamenti sperimentali per la conservazione attiva dei documenti sonori su nastro magnetico: individuazione delle curve di equalizzazione. In *Proceedings of the XXI Colloquium of Musical Informatics*, pages 114–121, Cagliari, September 2017.
11. O. Lartillot and P. Toiviainen. A matlab toolbox for musical feature extraction from audio. In *International Conference on Digital Audio Effects (DAFx-07)*, pages 237–244, September 2007.
12. Adam Berenzweig, Beth Logan, Daniel PW Ellis, and Brian Whitman. A large-scale evaluation of acoustic and subjective music-similarity measures. *Computer Music Journal*, 28(2):63–76, 2004.
13. John C Mallinson. Tutorial review of magnetic recording. *Proceedings of the IEEE*, 64(2):196–208, 1976.