

1. INTRODUCTION

Musical instrument classification is a fundamental task in Musical Information Retrieval (MIR), focusing on identifying the types of instruments present in a recording, such as violins, cellos, or pianos. This topic has been extensively studied in the literature, and for monophonic recordings (containing only one instrument), state-of-the-art models often achieve nearly 100% accuracy. However, few articles have addressed the issue of identifying individual instruments of the same type, let alone from the violin family specifically (for example, identifying violin A versus violin B). Only a handful of studies ([1], [2], [3]) have tackled this challenge, showing that cepstral coefficients provide sufficient discriminative information about individual violin timbre to enable effective classification. [1] found that long-term features could lead to a more stable classification. Nevertheless, these studies relied on limited datasets, where not all violins were played by multiple violinists. This raises an important question: do these algorithms recognize the instruments themselves, or simply the unique playing styles of the performers?

This paper aims to address this challenge by reviewing and comparing various approaches to identifying individual violins. In order to achieve this, two datasets were collected, comprising recordings of multiple violinists playing a set of violins. Various features and machine learning algorithms were then analyzed, ultimately providing retrospective guidelines for recording sessions — for example, identifying which musical excerpts are most effective for capturing the timbral differences between violins. The remainder of this paper is organized as follows: Section 2 describes the methodology, including data collection, feature extraction, data exploration, and classification using machine learning. The experimental results are presented and analyzed in Section 3. Finally, Section 4 provides concluding remarks and outlines potential directions for future work.

2. METHODOLOGY

This section outlines the methodology employed in this study, covering data collection, processing, and the machine learning algorithms applied.

A. DATASETS

Two datasets were collected for this study, each consisting of recordings where a group of violinists played a set of violins. The recording duration for each player and each violin in both datasets are presented in Figure 2.

i. Bilbao Dataset

During the Bilbao Project, thirteen violins were built in order to relate their material and geometrical characteristics with their tonal quality [4]. These violins have been played in 2019 by twenty-three professional violinists, each of them having recorded a G chromatic scale (3 octaves) on each violin and a short musical excerpt on a violin of their choice. Our dataset thus consists of 13×23 scales plus 1×23 musical excerpts. The recordings were made under the same conditions in a large rehearsal room at the Bilbao conservatory, keeping the distance between the player and the microphone (Zoom H4n Pro) constant and using a sample rate of 48kHz.



Figure 1: 3-octave G chromatic scale played by the participants of the recording.

Six of those thirteen Bilbao violins (violins number 1, 4, 5, 9, 11 and 13) were brought to the 2024 Villfavard Workshop and were recorded again. They were played freely by four new players in a small room, under rather different conditions than during the 2019 recordings.

ii. CNSM Dataset

In September 2024, we conducted an experiment at the Paris Conservatoire (CNSMDP) involving thirteen violinists and three violins. The participants were invited to freely explore each instrument before recording selected excerpts from the classical violin repertoire, including pieces by Bach (Allemande), Mozart (Concerto No. 3), Tchaikovsky (Concerto), Sibelius (Concerto), and Glazunov (Concerto).

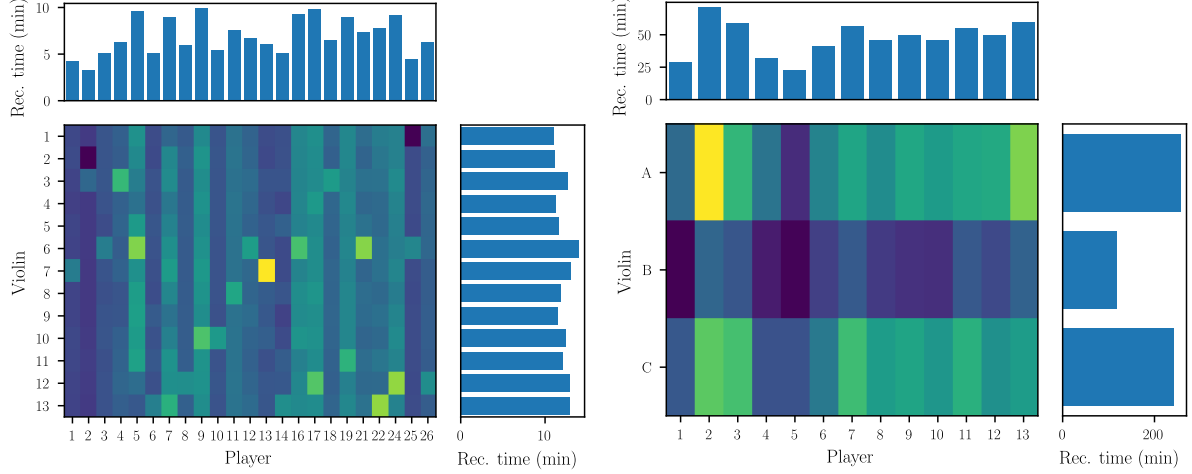


Figure 2: Recording time available with respect to players and with respect to violins

B. FEATURES

We limit the study to long-term audio descriptors, as we can observe that when violinists try new instruments they often give a global description of the sound they produce. The following features have been compared for the classification task :

i. Long Time Average Spectra (LTAS)

The Long Time Average Spectra (LTAS) of a recording is obtained by dividing the input signal into overlapping segments, then calculating the windowed DFT of each segment and finally averaging the power of those DFTs. These steps are shown in Figure 3 :

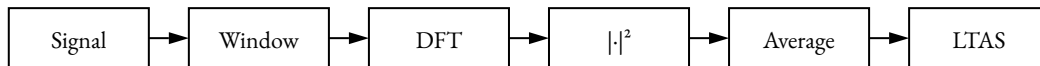


Figure 3: LTAS computation steps

LTAS has been used in [5] in order to compare the tonal quality of violins. More specifically, the sound of old Italians violins (Stradivari/Guarneri) and modern violins has been compared. The author concludes that differences between these two groups can be shown using LTAS

ii. Long-Term Cepstral Coefficients (LTCC)

LTCC have been introduced in [1] for Individual Instrument Identification. Their calculation is similar to that of MFCCs, except that a Mel-filterbank is not applied and that the final step is given by an Inverse Discrete Fourier Transform. These steps are shown in Figure 4 :

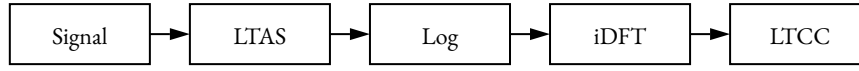


Figure 4: LTCC computation steps

iii. Mel-Frequency Cepstral Coefficients (MFCC)

MFCCs are obtained by mapping the frequencies of a spectrum onto a nonlinear mel-scale (a perceptual scale of pitches judged by listeners to be equal in distance from one another), taking the log, and then compute the DCT of the result. Here, instead of calculating the MFCCs on overlapping segments, we use a LTAS as our spectra as we want features with a long-term meaning. These steps are shown in Figure 5 :

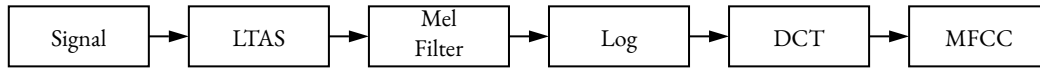


Figure 5: MFCC computation steps

MFCC are a set of features that has been extensively used for Automatic Speaker Recognition and for Instruments Classification ([6], [7]).

C. CLASSIFIERS

We compare the results of three popular classification algorithms : K-Nearest Neighbours, Support Vector Machines and Multilayer Perceptron. These three classifiers use different learning strategies and thus will give different results on our data.

i. K-Nearest Neighbors

K-Nearest Neighbours is a method that predicts the label of a new point based on the most frequent label among its K closest neighbors in the feature space. Since the prediction is made directly from the training data, this method is non-parametric (or non-generalizing), which is an advantage when the decision boundary is irregular.

ii. Support Vector Machines

Support Vector Machines is a supervised learning method using for classification. It works by finding an optimal hyperplane that maximizes the distance between each class in the training data. This algorithm is computationally expensive but generally has good generalisation properties.

iii. Multilayer Perceptron (MLP)

A Multilayer Perceptron is a type of neural network that maps inputs to outputs through multiple layers of interconnected neurons, using learned weights and biases to perform complex transformations. This method has demonstrated effectiveness in various domains, as the non-linear activation functions can model complex relationships in data while hidden layers can learn hierarchical representations from input data. However, MLPs are prone to over-fitting, especially on small datasets.

3. RESULTS & DISCUSSIONS

This section examines the influence of hyperparameters on the classification task and evaluates performance across different dataset subsets. Classifier scores were computed for a grid of hyperparameters. To analyze the effect of each hyperparameter, the mean and standard deviation of the scores were calculated while varying only that hyperparameter and keeping all others constant.

A. DATASET HYPERPARAMETERS

i. Influence of sample rate

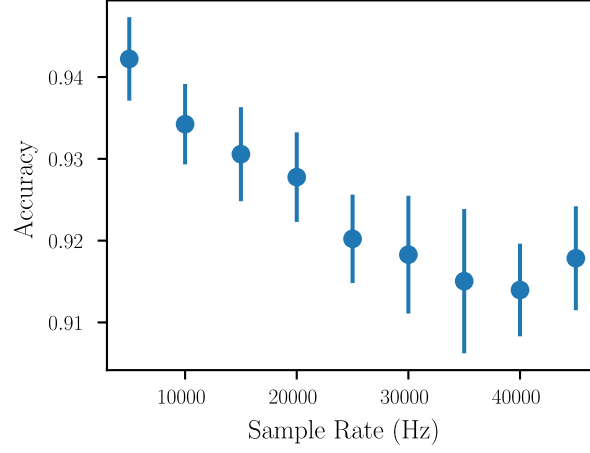


Figure 6: Influence of the sampling rate of the dataset on the accuracy of the algorithms

The results in Figure 6 indicate that high sampling rates are not necessary to achieve high accuracy. This suggests that the low-frequency components of the spectra provide sufficient discriminative information for accurate predictions. Consequently, using a relatively low sampling rate (e.g., 8000 Hz) is a more efficient choice, reducing storage requirements and computation time without compromising performance.

ii. Influence of sample duration

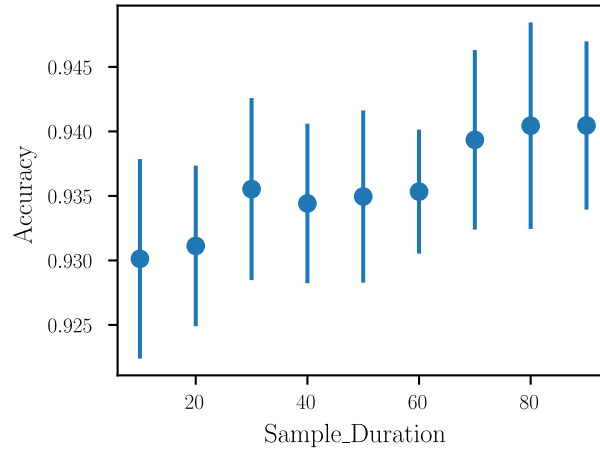


Figure 7: Influence of the sample duration on the accuracy of the algorithms

Analysis of Figure 7 reveals that longer sample durations lead to higher accuracy. This can be attributed to the use of long-term features, which require a wide range of played notes to effectively capture the violins' signature spectra. However, sample durations exceeding 40 seconds do not appear to provide additional improvements in accuracy.

B. FEATURES HYPERPARAMETERS

i. Influence of feature

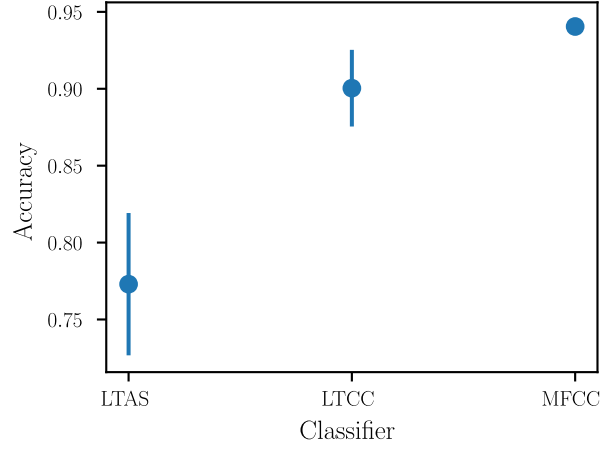


Figure 8: Influence of the sampling rate of the dataset on the accuracy of the algorithms

The results in Figure 8 indicates that MFCCs and LTAS seem to be better features to distinguish violins from their sound.

ii. Influence of the number of coefficients (for MFCCs and LTCCs)

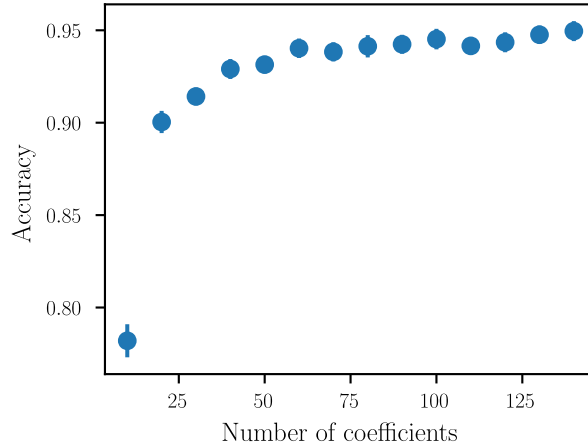


Figure 9: Influence of the number of coefficients kept in the calculation of LTCCs and MFCCS

In the computation of LTCCs and MFCCs, using fewer cepstral coefficients results in a smoother spectral representation, as only the first coefficients of the spectral transform are retained. The results in Figure 9 show that a high number of cepstral coefficients is not required to achieve good prediction performance. This indicates that even a smoothed representation of the spectra provides sufficient discriminative information for accurate classification.

C. CLASSIFIERS HYPERPARAMETERS

i. Influence of the classifier

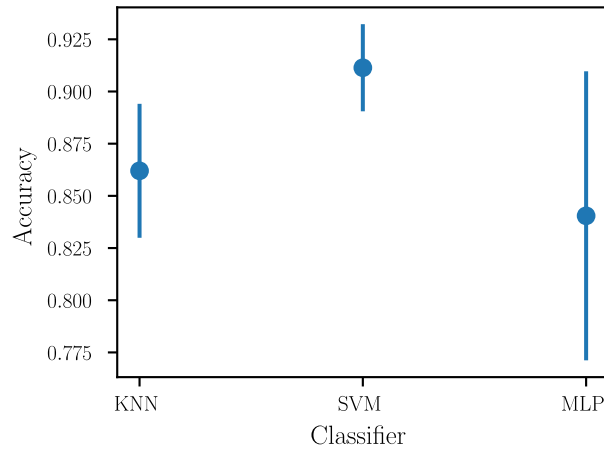


Figure 10: Influence of the sampling rate of the dataset on the accuracy of the algorithms

The results in Figure 10 show that all three classification algorithms achieve a similar accuracy, approximately 90%. Since these algorithms have different inherent disadvantages, this consistency suggests that the accuracy obtained is largely independent of the choice of classifier. The results indicate that KNNs and SVMs outperform MLPs. This suggests that the dataset size may be insufficient for MLPs to perform optimally. Additionally, the features appear to be sufficiently well-separated in the feature space, allowing simpler algorithms to achieve better results.

4. CONCLUSIONS

In the paper, we challenged the identification of individual violins based on their unique timbre, using machine learning techniques. By assessing the influence of multiple parameters on the classification accuracy, this study established that long-term MFCCs combined with a simple KNN are powerful enough to tackle this task. The low-frequency content is discriminative enough to allow a sample rate of only 8 kHz, providing that the each sample is long enough (30 seconds or longer). Future work will focus on correlating MFCCs with perceptual evaluations of violin timbre by musicians, providing insights into how objective metrics align with subjective listening experiences.

REFERENCES

- [1] E. Lukasik, "Long Term Cepstral Coefficients for Violin Identification," *Journal of The Audio Engineering Society*, May 2010, Accessed: Apr. 24, 2024. [Online]. Available: <https://www.semanticscholar.org/paper/Long-Term-Cepstral-Coefficients-for-Violin-Lukasik/1078124310e55d7f734e2cead68f0d1ad7ccf511>
 - [2] Q. Wang and C. Bao, "Individual Violin Recognition Method Combining Tonal and Nontonal Features," *Electronics*, vol. 9, no. 6, p. 950, Jun. 2020, doi: 10.3390/electronics9060950.
 - [3] M. Yokoyama and Y. Ishigaki, "Identification of Violin Timbre by Neural Network Using Acoustic Features," *Proceedings of Meetings on Acoustics*, vol. 49, no. 1, p. 35004, Dec. 2022, doi: 10.1121/2.0001659.
 - [4] C. Fritz, G. Stoppani, U. Igartua, R. J. Rico, A. Arroitauregi, and L. Artol, "The Bilbao Project : How Violin Makers Match Backs and Tops to Produce Particular Sorts of Violins," 2019.
 - [5] A. Buen, "Comparing the Sound of Golden Age and Modern Violins: Long-Time-Average Spectra," *Proceedings from BNAM*, Jul. 2005.
-

-
- [6] A. Eronen and A. Klapuri, "Musical Instrument Recognition Using Cepstral Coefficients and Temporal Features," in *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*, Jun. 2000, pp. II753–II756vol.2. doi: 10.1109/ICASSP.2000.859069.
- [7] J. D. Deng, C. Simmermacher, and S. Cranefield, "A Study on Feature Analysis for Musical Instrument Classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 2, pp. 429–438, Apr. 2008, doi: 10.1109/TSMCB.2007.913394.

5. APPENDIX