Individual violin sound identification using audio features and machine learning

Hugo Pauget Ballesteros, Claudia Fritz, Philippe Lalitte

June 2024

Few articles have addressed the issue of identifying individual instruments of the same type from their recordings. In this paper, we explore violin sound identification using audio features and machine learning algorithms. A review and a comparison between different approaches is made, which allow

First Author: First University. Second Author: Second University. We thank colleagues for helpful comments and discussions. This work was supported by a grant [grant number]; another grant [grant number]; and a foundation.

1. Introduction

Musical instruments classification is a Musical Information Retrieval (MIR) task which consists of determining the instruments present in a recording. This topic has been extensively studied in the literature, and for monophonic recordings (containing only one instrument), state-of-the-art models reach often almost 100%. However, few articles have addressed the issue of identifying individual instruments of the same type, let alone from the violin familiy specifically.

In Lukasik (2010),

The goal of this paper is to review and compare different ways of tackling this task, from data collection to data processing. Guidelines regarding recording sessions are discussed and a Long-Time version of MFCCs is introduced.

This paper is structured as follows: Section 2 presents the methodology of our experiment, describing data collection, features extraction, data exploration and finally classification using machine learning methods. Results of the experiments are discussed in Section 3. Finally, conlusions are drawn in Section 4, which also outlines possible future developments.

2. Methodology

2.1. Dataset

During the Bilbao Project, thirteen violins were built in order to relate their material and geometrical characteristics with their tonal quality (Fritz, Salvador, and Stoppani 2021). These violins have been played in 2019 by twenty-three professional violinists, each of them having recorded a scale on each violin and a short musical excerpt on a violin of their choice. 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 constant. Our dataset thus consists of 13×23 scales plus 1×23 musical excerpts.

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.

2.2. Features

The following features have been compared for the classification task :

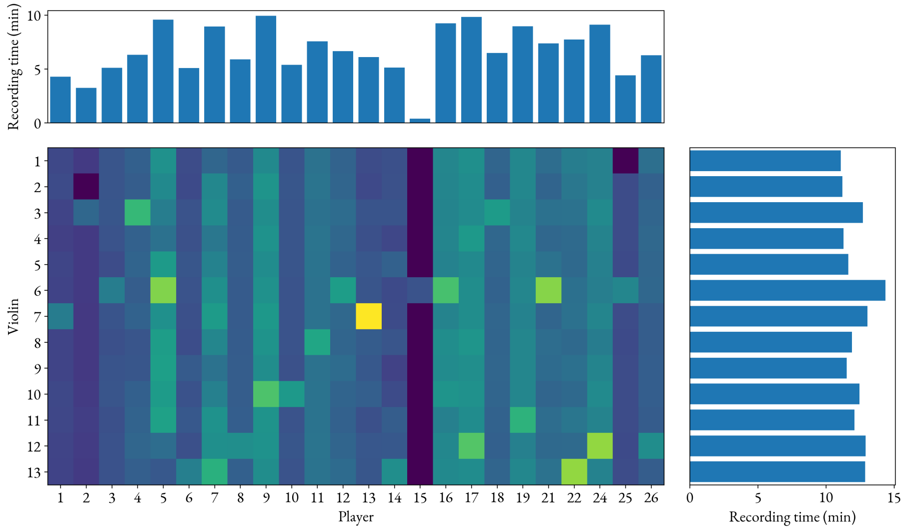


FIGURE 1. Recording time available with respect to players and with respect to violins

2.2.1. 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 :

LTAS has been used in Buen (2005) 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

2.2.2. Mel-Frenquency 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 :

MFCC are a set of features that has been extensively used for Automatic Speaker Recognition and for Instruments Classification.

2.2.3. Long-Term Cepstral Coefficients (LTCC)

LTCC have been introduced in Lukasik (2010) for Indivudial 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.

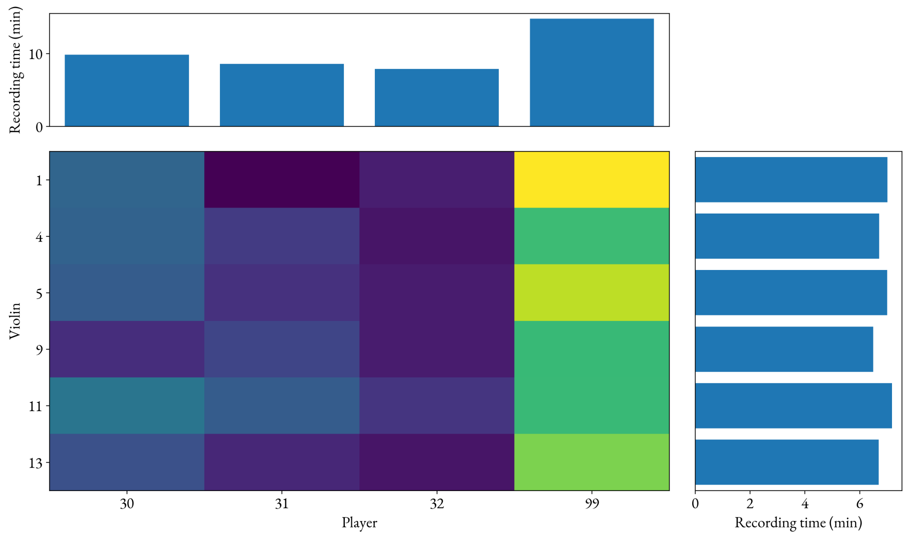
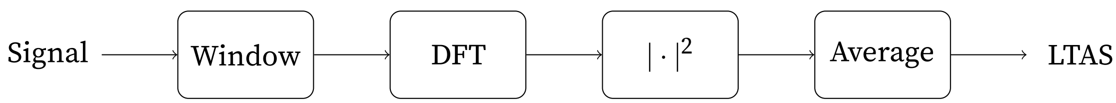


FIGURE 2. Recording time available with respect to players and with respect to violins



2.3. Data exploration

2.3.1. Variability of LTAS

2.3.2. Feature selection

2.4. Classification

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.

2.4.1. K-Nearest Neighbours

K-Nearest Neighbours is a method that finds the closest training points to a new test point and predicts its label from them. 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.

2.4.2. 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

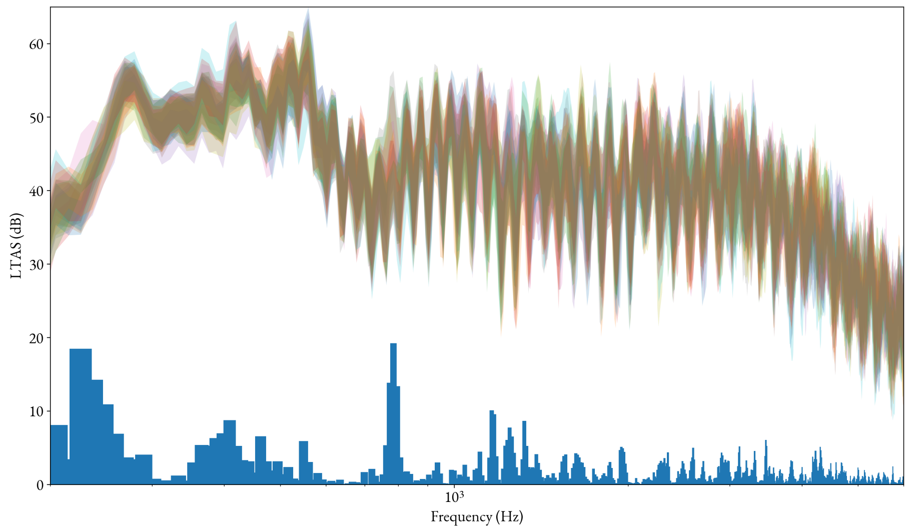
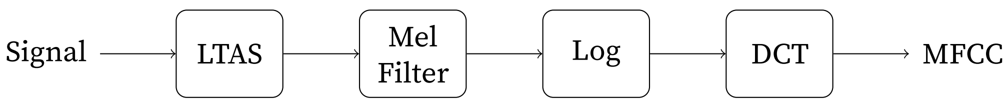


FIGURE 3. Standard Deviation of the LTAS of the 13 violins with respect to players



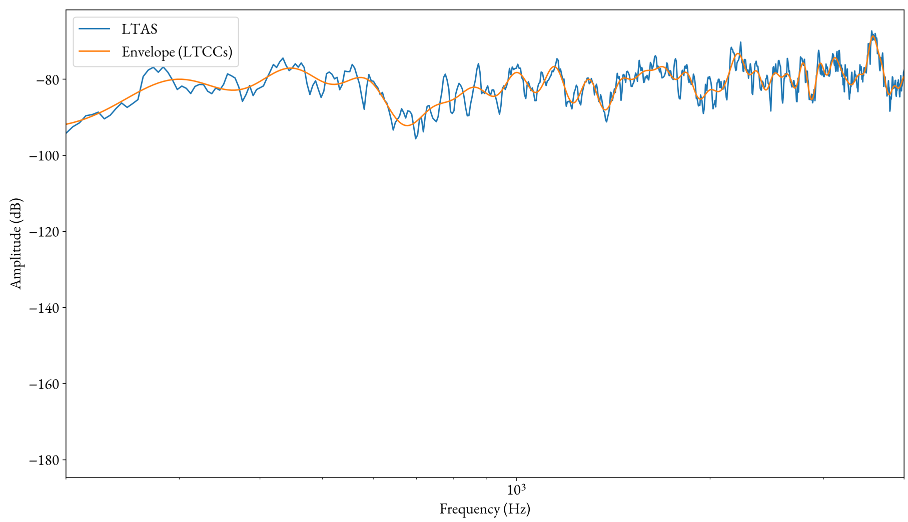
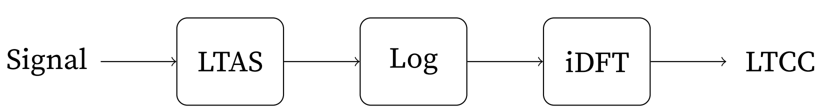


FIGURE 4. Standard Deviation of the LTAS of the 13 violins with respect to players



the training data. This algorithm is computationally expensive but generally has good generalisation properties.

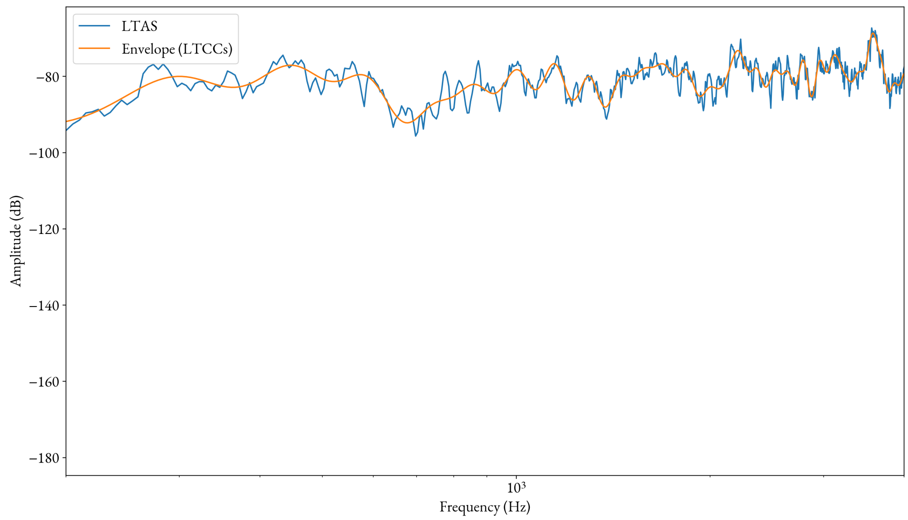


FIGURE 5. Standard Deviation of the LTAS of the 13 violins with respect to players

2.4.3. Multilayer Perceptron

Multilayer perceptron is a supervised learning method that learns a function $f : \mathbb{R}^n \to \mathbb{R}^o$ using training data. To do so, it uses layers of neurons as in Figure 6. Each neuron transforms the result from the previous layer by forming a weighted linear summation $w_0 + w_1x_1 + \cdots + w_nx_n$ followed by a non-linear activation function.

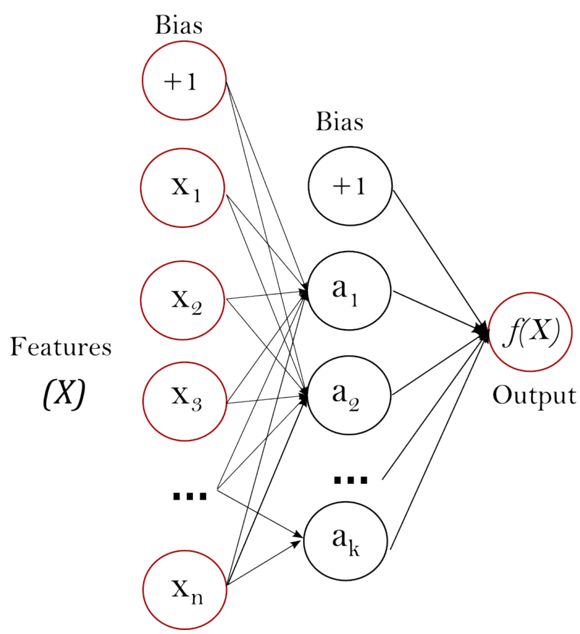


FIGURE 6. Multilayer Perceptron

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 overfitting, especially on small datasets.

3. Results

In this section we present the results of the classification task performed on different subsets of the dataset. We compare the results of each feature paired with each classifier. The results showed below are the one after fine-tuning the hyper-parameters, which can be hyper-parameters in the computation of the feature (sample rate, sample duration, ...) or hyper-paramters of the classification algorithm (k for K-NN, for instance).

TABLE 1. Results using scales as training data and muscial excertps as test data

Method	Feature	Train	Test
K-NN	LTAS	100%	37%
	LTCC	100%	100%
	MFCC	100%	100%
SVM	LTAS	100%	37%
	LTCC	100%	100%
	MFCC	100%	100%
MLP	LTAS	100%	37%
	LTCC	100%	100%
	MFCC	100%	100%

TABLE 2. Results using cross-validation on the entire dataset

Method	Feature	Train	Test
K-NN	LTAS	100%	37%
	LTCC	100%	100%
	MFCC	100%	100%
SVM	LTAS	100%	37%
	LTCC	100%	100%
	MFCC	100%	100%
MLP	LTAS	100%	37%
	LTCC	100%	100%
	MFCC	100%	100%

4. Conclusions

References

Buen, Anders. 2005. "COMPARING THE SOUND O F GOLDEN AGE AND MODERN VIOLINS: LONG-TIME-AVERAGE SPECTRA."

Fritz, Claudia, Víctor Salvador, and George Stoppani. 2021. "The Bilbao Project: Searching for Relationships between Sound and Playing Properties of Violins with Their Construction Parameters." In *Conference on Sound Perception*.

Lukasik, E. 2010. "Long Term Cepstral Coefficients for Violin Identification." *Journal of The Audio Engineering Society*.