Individual violin sound identification using audio features and machine learning

Hugo Pauget Ballesteros, Claudia Fritz, Philippe Lalitte

June 2024

This is the abstract. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

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.

In Zhao, Fazekas, and Sandler (2022)

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 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.

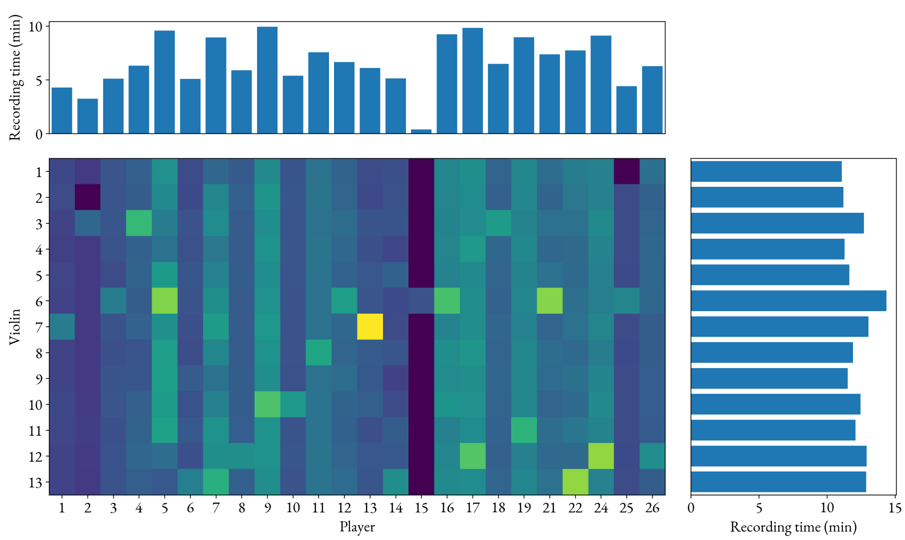


FIGURE 1. Recording time available with respect to players and with respect to violins

2.2. Features

The following features have been compared for the classification task :

2.2.1. Long-Term Average Spectra (LTAS)

The Average Power Spectral Density (PSD) 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 :

$$\text{Signal} \longrightarrow \boxed{\text{Window}} \longrightarrow \boxed{\text{DFT}} \longrightarrow \boxed{|\cdot|^2} \longrightarrow \boxed{\text{Average}} \longrightarrow \text{LTAS}$$

LTAS has been used [?] in order to compare the tonal quality of violins.

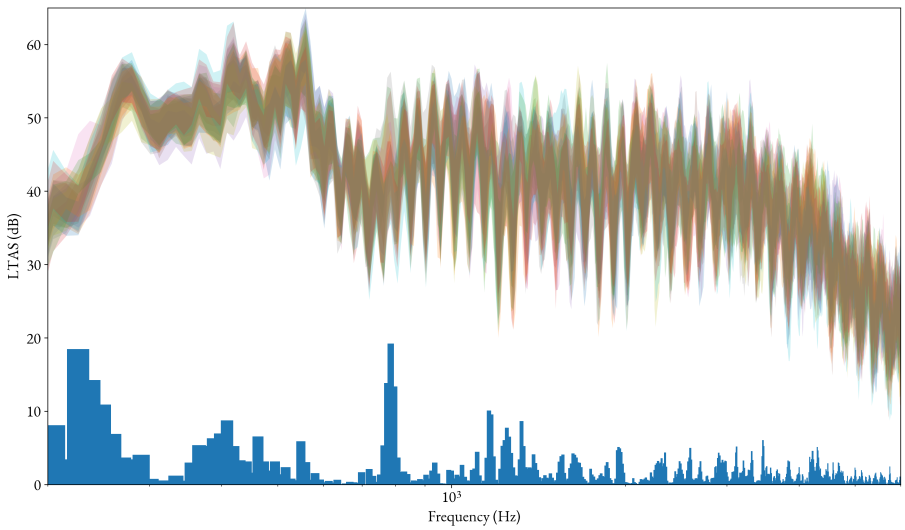


FIGURE 2. Standard Deviation of the LTAS of the 13 violins with respect to players

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 :

$$\text{Signal} \longrightarrow \boxed{\text{LTAS}} \longrightarrow \boxed{\begin{array}{c}\text{Mel}\\\text{Filter}\end{array}} \longrightarrow \boxed{\text{Log}} \longrightarrow \boxed{\text{DCT}} \longrightarrow \text{MFCC}$$

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

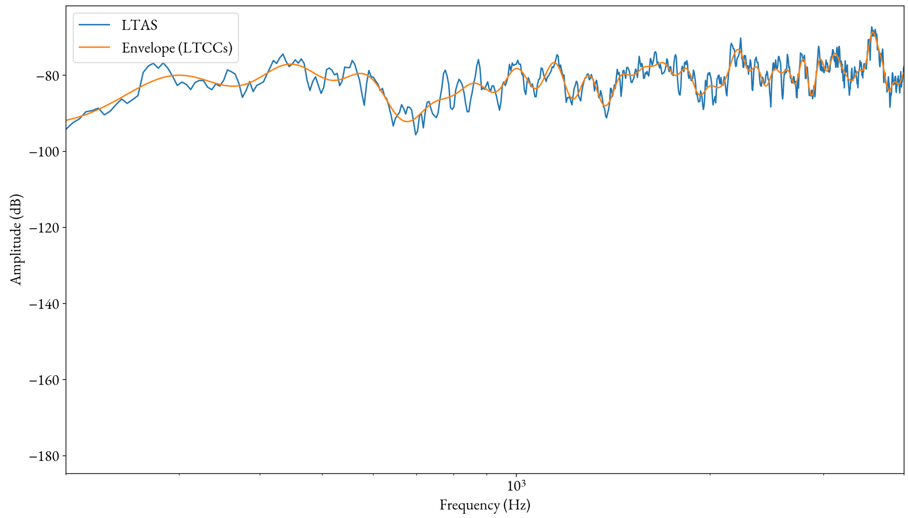
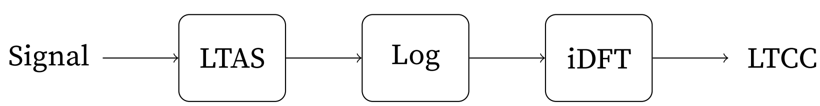


FIGURE 3. Standard Deviation of the LTAS of the 13 violins with respect to players

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.



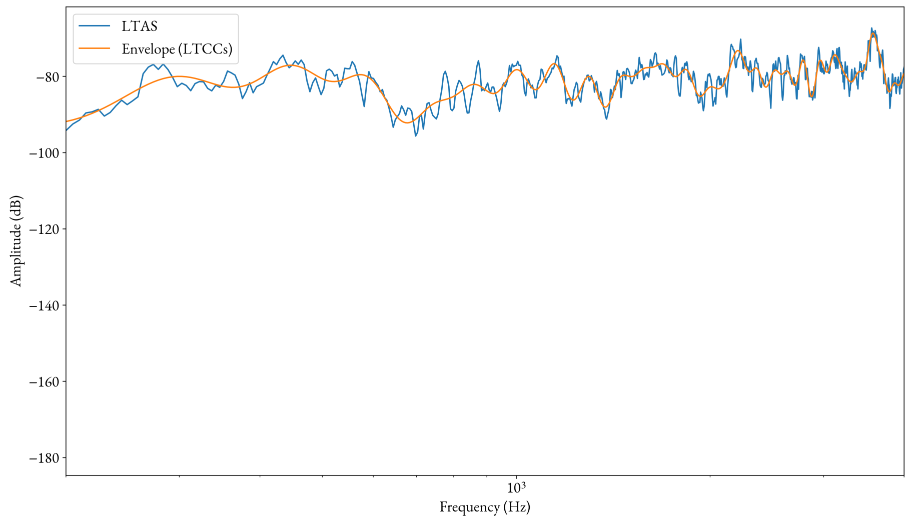


FIGURE 4. Standard Deviation of the LTAS of the 13 violins with respect to players

TABLE 1. the basic table

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%

2.3. Data exploration

2.3.1. Feature selection

2.4. Classification

2.4.1. K-Nearest Neighbours

2.4.2. Support Vector Machines

2.4.3. Multilayer Perceptron

3. Results

4. Conclusions

References

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*.

Zhao, Yudong, Gyorgy Fazekas, and Mark Sandler. 2022. "Violinist Identification Using Note-Level Timbre Feature Distributions." *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*: 601–605.