Feature-Based Timbral Characterization of Historical and Modern Violins

Francesco Setragno,[1†] *Massimiliano Zanoni,*[1] *Fabio Antonacci*[1] *Augusto Sarti*[1]

[1]Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB) - Politecnico di Milano - Piazza Leonardo da Vinci, 32, 20133 Milano

[†]francesco.setragno@polimi.it

ABSTRACT

Violin timbre is a very complex case of study. The sound properties that distinguish an historical violin from a modern one are still not clear. The purpose of this study is to understand what are these properties, by means of feature-based analysis. We extract audio features related to timbre and we exploit feature selection techniques in order to investigate what are the most characterizing ones. We compare different feature selection algorithms and we illustrate how we applied their outcome to a classification task with historical and modern instruments. Results show that the classification performance improves when using the selected features.

1. INTRODUCTION

Violin has been a subject of research for decades. This complex instrument has been examined from several points of view (vibro-acoustic, chemical, microscopic etc.). Among them, timbre is certainly very important.

One interesting aspect in the study of violin sound quality is the timbral characterization of historical instruments. Although the ability of discerning a historical violin from a modern one is not easy for human listeners [1], it is know that many factors influence the sound of historical instruments making them, in most cases, distinguishable from modern ones. Well known techniques in the literature for musical instrument sound characterization rely on feature-based analysis, that allows to extract a set of descriptors from the audio waveform able to capture specific aspects of the sound.

Feature-based methodologies have already been considered for the analysis of the violin tone. In [2] Łukasik et al. analyzed the dissimilarity factors of the timbre of various master violins, i.e. the features that enable to automatically identify the individual instruments in a pool. Feature-based timbral analysis that relies on feature-based representations is also applied to musical instrument identification. In [3] the authors use a set of multiscale Mel-Frequency Cepstral Coefficients for the pairwise discrimination of musical instruments. In [4] the author used Long Term Cepstral Coefficients to characterize the subtleties of violin sound, while in [5] several harmonic features are extracted from a collection of 53 violin recordings. The study also showed that the four strings exhibit different values for the same feature. Taking into account this result, in our work, we study each string separately.

Feature-based analysis is also used as the base for semantic description as introduced in [6], where a set of bipolar descriptors from natural language are modeled by means of large sets of acoustic cues extracted by 28 historical and modern violins.

In this paper we investigate the timbral properties of violins that help in discerning historical and modern violins by means of feature-based analysis. That is, if we look the problem by the information analysis view point, find the subset of features that best separate the two classes. Several feature selection algorithms are present in the literature. For this is paper we considered two feature selection and two feature ranking techniques. We then study the consistency of their results in order to provide with a final set of selected features. The selection and the understanding of discriminative proprieties, allows to implement applications able to automatically discern the two classes (historical and modern) of instruments by means of machine learning techniques. We used the feature as input for a binary classifier. In the results we show the performance of the classifier before and after the application of the feature selection procedure. We recorded 22 violins: 9 average modern violins and 13 historical violins from the collection of the Violin Museum in Cremona.

2. FEATURE ANALYSIS AND SELECTION

In this section we describe how the feature analysis and selection has been carried out.

Low-level features are objective descriptors that can be extracted from a signal by means of some signal processing techniques. Each feature captures a specific property of the audio signal. For this reason they are suitable for our purposes.

For this study, we extract low-level features that are well known in the Music Information Retrieval (MIR) field. In particular, we use the four statistical moments of the audio spectrum (centroid, spread, skewness and kurtosis), several features related to the shape of the spectrum (slope, rolloff, sharpness, smoothness, crest, flatness, irregularity), the spectral average deviation, the spectral entropy, several features related to harmonics distribution (tristimulus 1, 2 and 3, odd-even ratio) and the zero crossing rate (ZCR). Finally, we also extract two vectorial features: the Mel-Frequency Cepstral Coefficients (MFCC) and the Spectral Contrast (SP), which have been successfully used in the context of timbre analysis [7, 8]. The total number of features used is 108 (by considering each element of the MFCC and the Spectral Contrast as one feature). We refer to [9, 10, 11, 12, 13, 14] for a detailed explanation of these features. We use the Matlab MIR Toolbox [11] for extracting the features. Feature algorithms that were not provided by the toolbox are computed using ad-hoc developed software.

Each audio file is divided into 40 ms overlapping frames. Each frame is multiplied by a Hanning window and processed in order to extract the mentioned features. Therefore, for each frame we obtain a vector of length M, where M is the number of features.

To provide a more complete timbre characterization as possible, we extract a large set of audio features. This preliminary characterization is the input of the feature selection algorithm. Several algorithms are present in the literature based on specific selection measures. Tough many are effective, their results can be slightly different: even starting from the same input, the outputs of the different methods can be different. For this reason, in this study, we apply five algorithms, three for feature-selection and two for feature ranking, provided by Python's *sklearn* framework [15].

Feature selection algorithms output a subset of features according to specific criteria: SelectKBest and SelectPercentile select the K features and a given percentage of the features with the highest score, respectively, according to an ANOVA test. This test is useful because a feature which varies more from one group to another, rather than within the same group, represents a statistically relevant factor in discerning the two groups of instruments. False Positive Rate (FPR) selects the p-values below a threshold α based on a FPR test. Feature ranking algorithms produce a rank of features: the first we use is based on a Forest-Of-Trees, illustrated in [16], and the second is called Relieff [17].

We then compare the output of the selection methods in order to assess the consistence of their output. Features resulted to be selected by more algorithms are good candidates to compose the final set of features.

3. CLASSIFICATION

In order to prove that the selected features are actually relevant in the timbral characterization of historical violins, we setup a classifier to automatically discern modern and historical violins.

A classifier takes a vector of features as input and predicts the class of an instrument, according to a trained model. We use the well known Support Vector Classifier [18], that divides the feature spaces in such a way that the gap between each class is as wide as possible. We use a 3 order polynomial kernel in order to make the boundary nonlinear.

As reported in the next section, we record different sections ranging from open strings to musical excerpts. The classification task is performed separately for each section, since we want to test it for different type of content. In order to build the dataset, for each section the short frames are grouped into 5 seconds segments. Each segment contains the mean and the variance of every low-level feature and, for the frames belonging to the training set, a label which indicates if it is an historical or modern instrument (ground truth). In the training phase the extracted features are fed to the training algorithm, together with the correct label for each instrument (ground truth). In the test phase, the features are extracted from new audio segments and the trained model predicts the class of each segment.

The dataset is split as follows: 70% for the training set and 30% for the test set. The classification error is computed as the percentage of misclassified samples and this procedure is repeated five times. The classifier is trained with both the whole feature set and with the reduced one, that includes only the features selected by the feature selection algorithms. In this way we are able to investigate the impact of feature selection on the classification performance.

4. EXPERIMENTAL SETUP AND EVALUATION

In this section we provide the outcome of the feature selection procedure, comparing the result of each algorithm and showing the features that have been chosen most times. We then validate the set of selected feature in an historical versus modern automatic classification application.

Data Collection. In order to populate the dataset we recorded a set of instruments: 13 historical violins form the collection of the Violin Museum in Cremona and 9 modern violins from the violin making school Istituto Antonio Stradivari in Cremona.

The recordings were performed in a semi-anechoic room, using a measurement microphone always placed in the same position with respect to the instrument. In particular, the microphone was placed at a distance of approximately 40 cm, perpendicularly to the bridge. The audio was acquired with a sample rate of 44100 Hz. Since we did not want to include spatial information in this study, we acquired mono recordings. All the recordings were performed by the same professional musician that uses the same bow. For each instrument, the musician played the four open strings, a sequence of staccato notes on each string, a legato major scale from G4 to F#7 covering all the strings and 6 pieces of classical music including several styles and techniques. Therefore, each recording resulted in 15 sections: *1:Open G string; 2:Open D string; 3:Open A string; 4:Open E string; 5:Notes on G string; 6:Notes on D string; 7:Notes on A string; 8:Notes on E string; 9:Scale; 10:Excerpt 1; 11:Excerpt 2; 12:Excerpt 3; 13:Excerpt 4; 14:Excerpt 5; 15:Excerpt 6.*

Consistency between Feature Selection algorithms. In this section we investigate the consistency between the outputs of the adopted feature selection algorithms. We present the results by grouping the recorded sections into three subsets according to their musical content: open strings, single notes and scale/excerpts. We believed that the features that characterize the timbre of historical violins have different impact according to the nature of their content.

In order to compare and validate the results of the Relieff and Decision Tree ranking algorithms, we used the Spearman's rank correlation coefficient, that is able to compute how much two rankings are similar:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}, \quad (1)$$

Section	Spearman's coefficient
Open strings	0.80
Single notes	0.80
Scale and excerpts	0.69

Table 1. Spearman's rank correlation coefficient between the Relieff ranking and the forest-of-trees ranking, for each subset
.

Section	Fraction of common features
Open strings	0.80
Single notes	0.68
Scale and excerpts	0.70

Table 2. Fraction of features that all the three selection algorithms choose most times.

where d_i is the ranking difference for the i-th feature and n is the number of elements in the rankings.

We obtained good results on the correlation between the two ranking methods for almost every section. This means that the rankings were consistent. We reported them in table 1.

In order to study the consistency between the other methods, for each group of sections we took the 20 features that have been chosen most times by each algorithm. Then we intersected the three lists in order to find the features on which the three methods agreed. The result is showed in Table 2 as the number of common features divided by 20. The open strings, where the audio content is simpler and steadier, show the best consistency.

Results on Feature Selection. Features that are most discriminant for our task are those which are most frequently selected by the selection algorithms. Results of the selection is presented in Figure 1 where we report the most frequently selected features from the five methods.

It can be seen that the four statistical moments of the spectrum resulted discriminative, in particular for the sections that are characterized by steady sounds (open strings and single notes). Two MFCC coefficients, two Spectral Contrast coefficients and the first coefficient of the Tristimulus resulted important for the scale and the excerpts. This could be explained by the fact that, in presence of a more complex and variable audio content, more general features are needed, which do not represent just one specific aspect of the spectrum and better capture the complexity of the content.

As an example of discriminative ability of the features, in Figure 2 we show the bidimensional feature space represented by MFCC2/MFCC7.In this space, the samples relative to the open strings are arranged. Even using only a couple of features, a separation can be noticed between the two classes (in particular for G string and A string).

Classification results.

In this section we show the results of the historical versus modern violin classification in the case when all the features are used and in the case when only selected features are. The set of feature used in the latter case are the ones reported in 1. Table 3 summarizes the result by reporting the classification errors of the Support Vector Classifier. The best results are

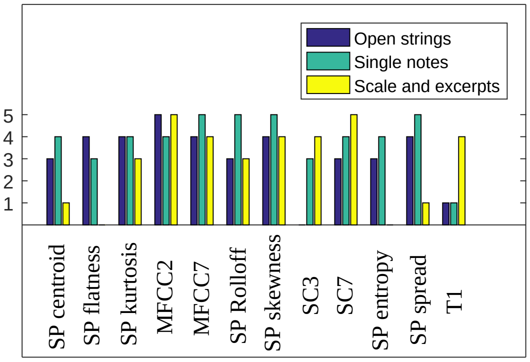

Figure 1. Number of algorithms that chose each feature for different sections. SC stands for Spectral Contrast, while T1 stands for Tristimulus (first coefficient)

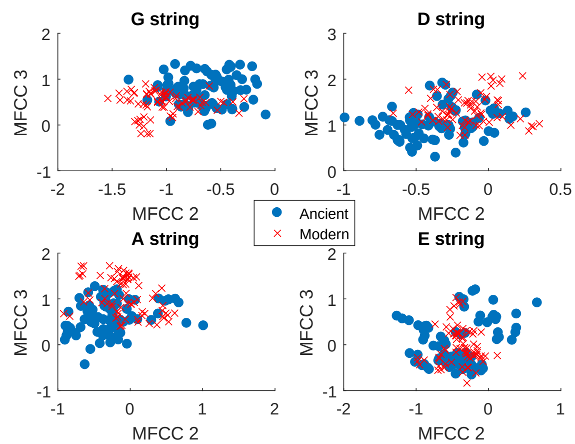

Figure 2. Values of the 2nd and 7th MFCCs for the open strings

achieved for the open strings. This could be due to the fact that the classification task is easier with steady notes, where there are not pauses and transients that affect the statistics of the low-level features. Table 4 shows the confusion matrix of the classification performed on the open strings.

In general, feature selection improves the classification results, meaning that the selected features are more relevant for this task.

5. CONCLUSIONS

In this work we investigated on the timbral aspects that are able to discern the sound of historical violins from the sound of contemporary ones, from a feature-based perspective. We applied different feature selection algorithms and we found out the most discriminative ones. In particular, the statistical moments of the spectrum and some MFCC and Spectral Contrast coefficients appeared to be important from this point of view. As we imagined, the relevance of such features depend

Section	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
AF	0.06	0.03	004	0.04	0.13	0.17	0.17	0.18	0.11	0.12	0.06	0.13	0.13	0.06	0.21
FS	0.02	0.01	0.03	0.02	0.20	0.15	0.18	0.16	0.10	0.11	0.10	0.10	0.18	0.03	0.19

Table 3. Classification error with the Support Vector Classifier, both with All features (AF) and with Feature Selection (FS)

	Historical	Modern
Historical	0.93	0.07
Modern	0.17	0.83

Table 4. Confusion matrix of the historical vs modern automatic classification.

on the audio content.

As a validation of the feature selection, we tested a classifier and proved that a classification model that is able to discern a modern violin from an old one can be implemented and its performances improve with feature selection.

This work constitutes the first step in understanding which are the sound properties that characterize the timbre of violins. The next step is to perform a qualitative and quantitative analysis on the audio features that appeared to be more relevant, and integrate this data with vibroacoustic measurements, in order to understand what are the structural factors that most impact these features. This will be addressed in future works.

REFERENCES

[1] C. Fritz, J. Curtin, J. Poitevineau, H. Borsarello, F.-C. Tao, and T. Ghasarossian, "Soloist evaluations of six old italian and six new violins," *Proceedings of the National Academy of Sciences*, vol. 111, no. 20, pp. 7224–7229, 2014.

[2] E. Lukasik and R. Susmaga, "Unsupervised machine learning methods in timbral violin characteristics visualization," in *Proc. Stockholm Music Acoustics Conference SMAC*, vol. 3, 2003, pp. 83–86.

[3] B. L. Sturm, M. Morvidone, and L. Daudet, "Musical instrument identification using multiscale mel-frequency cepstral coefficients," in *Signal Processing Conference, 2010 18th European*. IEEE, 2010, pp. 477–481.

[4] E. Lukasik, "Long term cepstral coefficients for violin identification," in *In proceedings of the Audio Engineering Society Convention 128 (AES128)*, 2010.

[5] ——, "Matching violins in terms of timbral features," *Archives of Acoustics*, vol. 31, no. 4, p. 227, 2006.

[6] M. Zanoni, F. Setragno, F. Antonacci, A. Sarti, G. Fazekas, and M. B. Sandler, "Training-based semantic descriptors modeling for violin quality sound characterization," in *Audio Engineering Society Convention 138*. Audio Engineering Society, 2015.

[7] H. Terasawa, M. Slaney, and J. Berger, "Perceptual distance in timbre space." Georgia Institute of Technology, 2005.

[8] D. Jang, M. Jin, and C. D. Yoo, "Music genre classification using novel features and a weighted voting method," in *2008 IEEE International Conference on Multimedia and Expo*. IEEE, 2008, pp. 1377–1380.

[9] T. S. H.G. Kim, N. Moreau, *MPEG-7 Audio and Beyond. Audio Content Indexing and Retrieval*. John Wiley & Sons Ltd, 2005.

[10] D. N. Jiang, L. Lu, H. J. Zhang, J. H. Tao, and L. H. Cai, "Music type classification by spectral contrast features," in *Proceedings of IEEE International Conference Multimedia Expo*, 2002.

[11] O. Lartillot and P. Toiviainen, "Mir in matlab (ii): A toolbox for musical feature extraction from audio," in *2007 International Society for Music Information Retrieval conference (ISMIR)*, 2007.

[12] K. Jensen, "Timbre models of musical sounds, tech. rep. rapport 99/7," University of Copenhagen, Tech. Rep., 1999.

[13] R. Plomp and W. J. M. Levelt, "Tonal consonance and critical bandwidth," *Journal of the Acoustical Society of America*, vol. vol. 38, pp. pp. 548–560, 1965.

[14] P. Juslin, "Cue utilization in communication of emotion in music performance: relating performance to perception." *Journal of Experimental Psychology: Human Perception and Performance*, vol. vol. 26, no. no. 6, pp. pp. 1797–1813, 2000.

[15] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler *et al.*, "Api design for machine learning software: experiences from the scikit-learn project," *arXiv preprint arXiv:1309.0238*, 2013.

[16] P. Geurts, D. Ernst, , and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63(1), pp. 3–42, 2006.

[17] K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm," in *AAAI*, vol. 2, 1992, pp. 129–134.

[18] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.