VIOLINIST IDENTIFICATION USING NOTE-LEVEL TIMBRE FEATURE DISTRIBUTIONS

Yudong Zhao,[1] *György Fazekas,*[1] *Mark Sandler,*[1]

[1] Centre for Digital Music, School of Electronic Engineering and Computer Science
Queen Mary University of London, Mile End Road, London E1 4NS, UK

ABSTRACT

Modelling musical performers' individual playing styles based on audio features is important for music education, music expression analysis and music generation. In violin performance, the perception of playing styles are mainly affected by the characteristic musical timbre, which is mostly determined by performers, instruments and recording conditions. To verify if timbre features can describe a performer's style adequately, we examine a violinist identification method based on note-level timbre feature distributions. We first apply it using solo datasets to recognise professional violinists, then use it to identify master players from commercial concerto recordings. The results show that the designed features and method work very well for both datasets. The identification accuracy with the solo dataset using MFCCs and spectral contrast features are 0.94 and 0.91 respectively. Significantly lower but promising results are reported with the concerto dataset. Results suggest that the selected timbre features can model performers' individual playing reasonably objectively, regardless of the instrument they play.

Index Terms— violinist identification, timbre feature distribution

1. INTRODUCTION

Musical structures established by composers and their interpretation by performers are two key factors that impact expressive music performance. The diversity of musical expression mostly depends on the characterisation or individual interpretation by different performers. In violin performance, although the performer's individual style is influenced by left-hand playing techniques (such as vibrato), it is strongly affected by the bowing gestures such as bow velocity, force, acceleration or bow-bridge distance. In a previous study [1], bowing data were acquired and measured using a hardware systems. However, the use of expensive sensing systems and complex setups are often intrusive in practice.Timbre features extracted from audio are capable of characterising violin bowing parameters to a good extent [2], while timbre variations are characteristic of a performer's individual preference and personal style [3].

There are many previous studies focusing on musical expression analysis and performer classification. Stamatatos and Widmer [4] proposed a set of features such as time deviation and melody lead [5] that capture aspects of pianists' individual style. Saunders et al. [6] have applied string kernels to the problem of recognising famous pianists by style. Ramirez et al. [7] developed a machine learning approach to identify Jazz saxophonists by analysing the pitch, timing, amplitude and timbre of individual notes.

There are also prior works on violin expression analysis and violinist classification. Li et al. [8] developed a dataset containing 11 expressive characteristics, then selected duration, dynamics and vibrato features to classify expressions using Support Vector Mechines (SVMs). Ramirez et al. [9] built a Celtic violinist classifier using machine learning method. Molina et al. [10] proposed an approach for identifying violinists on monophonic audio recordings using a musical trend-based model. Shih et al. [11] extracted articulation and energy features to compare different playing styles of Heifetz and Oistrakh. The authors in [12] presents a leading violinist identification method based on vibrato features and onset time deviations, while similar statistical methods in the context of piano performance proved valuable in [13].

To best of our knowledge, most previous works attempted violinist identification using features of pitch, timing, energy or vibrato, but the variation of timbre features has never been used to distinguish violinists. In this paper, we propose note-level timbre feature distributions to model performers' playing style, and present a method to identify violinists using these distributions. The flow chart of our approach is shown in Figure 1. We firstly construct datasets from solo musical scale recordings and concerto collections separately, then annotate onset time of each note manually after audio data pre-processing, which is introduced in Section 2. Next, we extract note-level timbre features and standardise them, followed by representing timbre variation characteristics for each performer by calculating global histogram distributions of each feature using all notes played by each performer. This is presented in Section 3. The details of the experiment setup as well as the results are discussed in Section 4. Finally, Section 5 provides conclusions and outlines possible future developments.

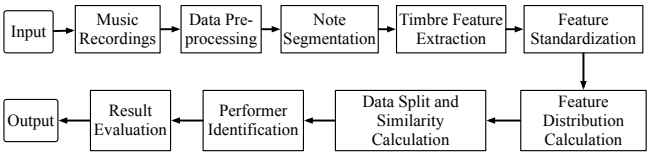


Fig. 1. Flow chart of the proposed methodology.

2. DATASETS

2.1. Solo violin dataset

During the European Bilbao project[1], thirteen new (white) violins were designed and built and evaluated within a free categorisation task by ten professional violinists. In this study, participants were invited to play a scale on each of the violins. The recordings were in a large rehearsal room at the Bilbao conservatory under the same conditions, keeping the position of the player and the microphone constant. We selected a group of ten players for our dataset, which consists of 10×13 musical scales in total. Each scale contains around 37 notes.

Since we aim to analyse note-level timbre features among different performers, the onset time label of each note needs to be accurate. Although there are many automatic onset detectors, the accuracy on violin recordings is not sufficient for our purposes, therefore we labeled onset times manually.

2.2. Concerto dataset

To investigate expressive performances of master players, we created a dataset of violin concerto pieces due to the genre's focus on the solo instrument. We selected five concertos written by five well-known composers listed in Table 1. These pieces have all been performed by nine violinists: Jascha Heifetz, Anne Sophie Mutter, David Oistrakh, Itzhak Perlman, Pinchas Zukerman, Isaac Stern, Salvatore Accardo, Yehudi Menuhin and Maxim Vengerov, who are all leading master violinists.

To extract note-level timbre features, we first segment the original music into several clips, then select at least 2 clips from each movement and annotate onset times for each note in the clips. The process of segmenting and selecting original concerto movements and the onset annotation method are discussed in [12]. Details of the recordings and the number of labeled notes in each movement are provided in Table 1.

3. METHODOLOGY

In this section, we present the method of violinist identification based on timbre feature distributions. To reduce possible differences in timbre characteristics due to varying recording conditions in both datasets, the data was pre-processed first.

Table 1. Concerto note segmentation dataset ('annotations' refers to the number of note annotations in each movement).

Composer	Concerto Name	Movement	annotations
L. V. Beethoven	Violin Concerto in D major, Op.61	I	664
		II	239
		III	352
J. Brahms	Violin Concerto in D major, Op.77	I	262
		II	157
		III	193
F. Mendelssohn	Violin Concerto in E minor, Op.64	I	204
		II	201
		III	235
P. I. Tchaikovsky	Violin Concerto in D major, Op.35	I	225
		II	177
		III	148
J. Sibelius	Violin Concerto in D minor, Op.47	I	233
		II	200
		III	186

Details of this process are introduced in Section 3.1. Timbre feature extraction and violinist identification methods are presented in subsequent sections respectively.

3.1. Data pre-processing

As outlined in Section 2, in the concerto dataset, recording conditions including concert hall or studio environment, musical instrument, microphone configuration and types as well as the accompanying philharmonic orchestra are different. This will influence timbre features and the identification performance. In the solo violin dataset, although all performers used same studio and microphone, it is hard to ensure their distance and direction from the microphone are kept constant. Therefore, to make the extracted features more comparable among performers, we first remove silent regions in each music clip, then loudness normalisation is applied using the EBU standard [14]. All steps were completed in Audacity[2].

3.2. Timbre feature extraction

We selected features that are either commonly used in the literature in related tasks, or have been validated in the context of violin bowing technique recognition in [2]. Six timbre related features are considered. One feature represents spectral moments (Spectral Centroid), three features describe primarily the shape of the spectrum (Mel-Frequency Cepstral Coefficients, Spectral Bandwidth [15] and Spectral Contrast [16]) and further two are temporal features (RMS energy and Zero-crossing rate). The reader is referred to [17, 18, 19, 20] for a detailed discussion of these features. The segmented notes in both datasets are divided into short overlapping frames (f_s=44.1 kHz, frame length = 2048, hop size = 512) and for each frame all features are extracted.

[1] https://www.bele.es/en/bilbao-project-introduction

[2] https://www.audacityteam.org/

3.3. Feature representation and modeling

There are many factors that influence timbre features including instruments, performers, and recording conditions. Furthermore, features such as spectral centroid or MFCCs are also influenced by the note pitch. Therefore, raw features are not directly comparable between performers. In this research, we assume the variation of timbre features within a note are mostly affected by performer's individual playing, and the distribution of timbre variations could model performers' characteristic performance.

3.3.1. Feature standardisation

We calculate the z-score of each feature vector at the note-level, which means standardising features by removing the mean and scaling to unit variance, i.e., the standard z-score of each sample x in the feature vector is calculated using Eq. 1:

$$z = \frac{x - \mu}{\sigma}, \tag{1}$$

where μ is the mean value of the feature vector, and σ is the standard deviation.

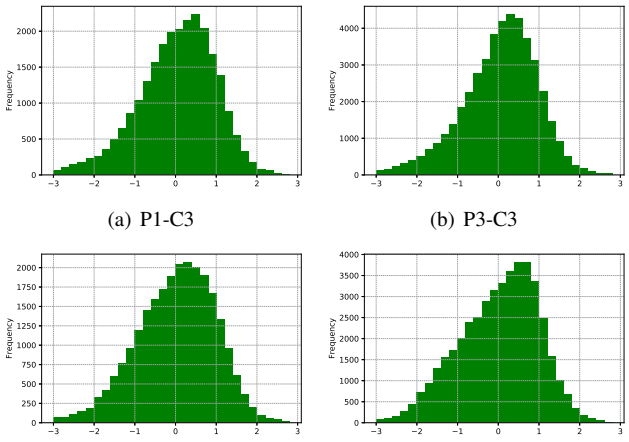
3.3.2. Feature distributions

We use histograms to calculate feature distributions of all notes played by each performer assuming these provide compact representations of the violinists' style, which we can use later for identification. Particularly, for multi-dimensional features, we use 2D histograms to model such data distributions.

Figure 2 shows the global distribution of the 3rd coefficient of MFCCs (c3) for four performers in solo dataset, where the x axis means the range of standardised feature, and the y axis presents the frequency. We abbreviate 'Perfomer1' as 'P1', and the same abbreviation is applied to all ten performers. The shape of such distributions are different, as demonstrated in the figure, which provides the basis for our hypothesis that these features are capable of characterising individual differences in performance style well. The sharpness, position of highest bar, and slope are different among the observed distributions. Based on similar observations across different performers and features, we assume that such features indeed reflect an important aspect of the performer's individual timbre characteristics.

3.4. Violinist identification

In order to quantify these differences, we calculate the similarity of distributions of each feature for all performers using the Kullback-Leibler (KL) divergence [21] $D_{KL}(P||Q)$ shown in Eq.2. This corresponds to the likelihood ratio between two distributions and tells us how well the probability



(a) P1-C3 (b) P3-C3

(c) P4-C3 (d) P5-C3

Fig. 2. Distribution of four performers' standardised MFCC (c3) feature in solo dataset.

distribution Q approximates the probability distribution P by computing the cross-entropy minus the entropy.

$$D_{KL}(P \,||\, Q) = \sum_i P(i) \log\left(\frac{P(i)}{Q(i)}\right) \tag{2}$$

For classification, the KL divergence is calculated between each timbre feature distributions of an unknown performer and every known performer in the dataset. Minimum divergence identifies the unknown performer. Classification experiments using this approach are presented in the Section 4.

4. EXPERIMENTS AND RESULTS

In this section, we first introduce three baseline methods and identify violinist in our two datasets. Then apply the proposed novel violinist identification method on the solo violin dataset and concerto dataset separately, investigating how it performs in identifying violin players under different conditions.

For each experiment, we test the proposed identification method using *Leave-One-Group-Out Cross Validation* (LOOCV) and show the classification result (F-measure) for all performers in the dataset. The results using each feature are also discussed and shown separately.

4.1. Baseline methods

We use three classifiers including K-Nearest Neighbour (KNN), Gaussian Mixture Model with KL divergence (GMM-KL) and Gaussian Mixture Model with Universal Background Model (GMM-UBM) as baseline, and MFCCs are set as input feature because these perform best in previous works as well as in our proposed method. These baseline methods are

Table 2. Violinist identification results using two datasets.

Method	Feature	Concerto dataset			Solo dataset		
		Precision	Recall	F-score	Precision	Recall	F-score
KNN	MFCCs	0.253	0.216	0.234	0.469	0.435	0.451
GMM-KL	MFCCs	0.279	0.243	0.250	0.588	0.542	0.553
GMM-UBM	MFCCs	0.323	0.316	0.319	0.789	0.763	0.775
Proposed Method	Spectral Centroid	0.235	0.236	0.235	0.459	0.438	0.439
	RMS	0.170	0.167	0.165	0.789	0.777	0.781
	Spectral Bandwidth	0.179	0.194	0.170	0.370	0.369	0.365
	Zero-crossing rate	0.137	0.135	0.136	0.243	0.246	0.235
	Spectral Contrast	0.324	0.283	0.302	0.918	0.908	0.908
	MFCCs	**0.341**	**0.333**	**0.326**	**0.941**	**0.938**	**0.937**

used for violinist identification [10], music similarity estimation [22], and violin classification [23], therefore we adopt them as baseline to identify violinists for both datasets, the details of data split and experiment setup are kept the same as in the experiments in Section 4.2 and Section 4.3.

4.2. Violinist identification using the solo dataset

In this experiment, we first select a performer as test performer, then designate one musical scale that was played with a certain violin from that performer as test data. Other musical pieces played with the selected violin from other performers are left out, whereas the remaining pieces from all performers (including the test performer) are placed in training set.

We compute the KL divergence between each feature's distribution from test performer and the same features for every performer in the training data. The similarity results for timbre characteristics based on six features can be separately obtained between the test performer and every performer in the training set. The smaller the KL divergence the greater the similarity, therefore we treat the performer that corresponds to the minimum value as the identified performer with each feature. This is effectively a nearest neighbour classification scheme.

Table 2 shows the F-measure result of violinist identification using each feature distribution separately. MFCCs work best among all features, which suggests the feature has good discrimination power on performers. The confusion matrix is shown in Figure 3 corroborating our observation.

4.3. Master violinist identification using concertos

Next, we assess the same method tested on scales to violin notes extracted from concerto recordings. In order to avoid overlapping musical segments between the training and test sets, we use movement-level LOOCV in this experiment. We designate all annotated notes from one concerto played by all 9 performers as the test set, while the remaining recordings are placed into the training set, thus 15-fold cross validation is applied in this task. We use same KL divergence calculation to identify violinist. The results for timbre characteristics based on six timbre features are also shown in Table 2.

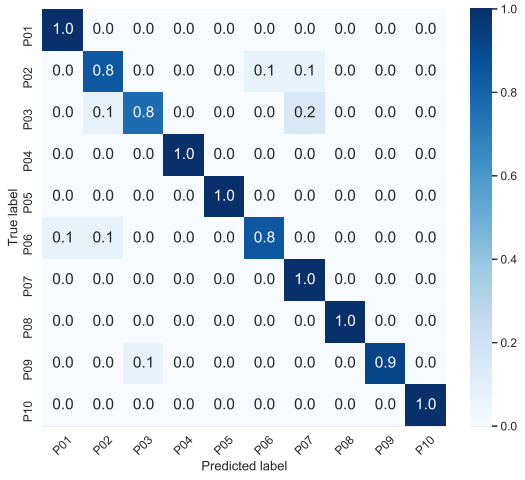


Fig. 3. Normalised confusion matrix for violinist identification using standardised MFCC feature distributions.

Similarly to the results from solo dataset, MFCC and Spectral Contrast are the best two features for identifying violinists, which confirms these features are helpful to identify performers.

5. DISCUSSION AND CONCLUSION

Given the results obtained in Table 2, we conclude that the proposed method works very well on solo violin dataset (all performers can be identified correctly) especially using the MFCC features. Similarly, timbre features are helpful to recognise violinists from concerto dataset although the results are somewhat less convincing. The results indicate that the distribution of note-level standardised timbre feature can reasonably capture and model the violinists' individual playing style.In the future, a larger dataset and additional features such as vibrato [24] may yield more robust result. To avoid the influence of accompaniment, we may apply source separation to isolate the violin performance.

6. REFERENCES

[1] Esteban Maestre, Panagiotis Papiotis, Marco Marchini, Quim Llimona, Oscar Mayor, Alfonso Pérez, and Marcelo M Wanderley, "Enriched multimodal representations of music performances: Online access and visualization," *IEEE Multimedia*, vol. 24, no. 1, 2017.

[2] Alfonso Perez-Carrillo, "Violin timbre navigator: Real-time visual feedback of violin bowing based on audio analysis and machine learning," in *International Conference on Multimedia Modeling*. Springer, 2019, pp. 182–193.

[3] Knut Guettler, "On the creation of the Helmholtz motion in bowed strings," *Acta Acustica united with Acustica*, vol. 88, no. 6, pp. 970–985, 2002.

[4] Efstathios Stamatatos and Gerhard Widmer, "Automatic identification of music performers with learning ensembles," *Artificial Intelligence*, vol. 165, no. 1, pp. 37–56, 2005.

[5] Werner Goebl, "Skilled piano performance: Melody lead caused by dynamic differentiation," in *Proc. 6th Int. Conf. on Music Perception and Cognition*, 2000.

[6] Craig Saunders, David R Hardoon, John Shawe-Taylor, and Gerhard Widmer, "Using string kernels to identify famous performers from their playing style," in *European Conference on Machine Learning*. Springer, 2004, pp. 384–395.

[7] Rafael Ramirez, Esteban Maestre, Antonio Pertusa, Emilia Gomez, and Xavier Serra, "Performance-based interpreter identification in saxophone audio recordings," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 3, pp. 356–364, 2007.

[8] Pei-Ching Li, Li Su, Yi-hsuan Yang, Alvin WY Su, et al., "Analysis of expressive musical terms in violin using score-informed and expression-based audio features.," in *ISMIR*, 2015, pp. 809–815.

[9] Rafael Ramirez, Esteban Maestre, Alfonso Perez, and Xavier Serra, "Automatic performer identification in celtic violin audio recordings," *Journal of New Music Research*, vol. 40, no. 2, pp. 165–174, 2011.

[10] Miguel Molina-Solana, Josep Lluís Arcos, and Emilia Gomez, "Identifying violin performers by their expressive trends," *Intelligent Data Analysis*, vol. 14, no. 5, pp. 555–571, 2010.

[11] Chi-Ching Shih, Pei-Ching Li, Yi-Ju Lin, AWY Su, L Su, and YH Yang, "Analysis and synthesis of the violin playing styles of Heifetz and Oistrakh," in *Proc. Int. Conf. Digital Audio Effects*, 2017.

[12] Yudong Zhao, György Fazekas, and Mark Sandler, "Identifying master violinists using note-level audio features," in *Sound and Music Computing Conference*, 2020.

[13] Syed Rifat Mahmud Rafee, G. Fazekas, and G. A. Wiggins, "Performer identification from symbolic representation of music using statistical models," in *International Computer Music Conference (ICMC), Santiago, Chile, July 25-31*, 2021.

[14] R EBU, "128, loudness normalisation and permitted maximum level of audio signals," *EBU Recommendation, Geneva*, 2014.

[15] Anssi Klapuri and Manuel Davy, "Signal processing methods for music transcription," 2007.

[16] Dan-Ning Jiang, Lie Lu, Hong-Jiang Zhang, Jian-Hua Tao, and Lian-Hong Cai, "Music type classification by spectral contrast feature," in *Proceedings. IEEE International Conference on Multimedia and Expo*. IEEE, 2002, vol. 1, pp. 113–116.

[17] Hyoung-Gook Kim, Nicolas Moreau, and Thomas Sikora, *MPEG-7 audio and beyond: Audio content indexing and retrieval*, John Wiley & Sons, 2006.

[18] Kristoffer Jensen, *Timbre models of musical sounds*, Ph.D. thesis, Department of Computer Science, University of Copenhagen Copenhagen, 1999.

[19] Reinier Plomp and Willem Johannes Maria Levelt, "Tonal consonance and critical bandwidth," *The journal of the Acoustical Society of America*, vol. 38, no. 4, pp. 548–560, 1965.

[20] Patrik N Juslin, "Cue utilization in communication of emotion in music performance: Relating performance to perception.," *Journal of Experimental Psychology: Human perception and performance*, vol. 26, no. 6, pp. 1797, 2000.

[21] Solomon Kullback and Richard A Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[22] Michael I Mandel and Daniel PW Ellis, "Song-level features and support vector machines for music classification," 2005.

[23] Qi Wang and Changchun Bao, "Individual violin recognition method combining tonal and nontonal features," *Electronics*, vol. 9, no. 6, pp. 950, 2020.

[24] Yudong Zhao, Changhong Wang, György Fazekas, Emmnouil Benetos, and Mark Sandler, "Violinist identification based on vibrato features," in *29th European Signal Processing Conference (EUSIPCO)*, 2021.