

# Violinist identification based on vibrato features

Yudong Zhao, Changhong Wang, György Fazekas, Emmanouil Benetos, Mark Sandler

Centre for Digital Music, Queen Mary University of London, UK

Email: {yudong.zhao; changhong.wang; g.fazekas; emmanouil.benetos; mark.sandler}@qmul.ac.uk

**Abstract**—Identifying performers from polyphonic music is a challenging task in music information retrieval. As a ubiquitous expressive element in violin music, vibrato contains important information about the performers’ interpretation. This paper proposes to use vibrato features for identifying violinists from commercial orchestral recordings. We present and compare two systems, which take the same note-level melodies as input while using different vibrato feature extractors and classification schemes. One system calculates vibrato features according to vibrato definition, models the feature distribution using histograms, and classifies performers based on the distribution similarity. The other system uses the adaptive wavelet scattering which contains vibrato information and identifies violinists with a machine learning classifier. We report accuracy improvement of 19.8% and 17.8%, respectively, over a random baseline on piece-level evaluation. This suggests that vibrato notes in polyphonic music are useful for master violinist identification.

**Index Terms**—Music signal processing, vibrato feature, violinist identification, statistical modelling, wavelet scattering

## I. INTRODUCTION

Musical structures established by composers and their interpretation by performers are two key factors that impact music performance [21]. Modelling master performers’ individual playing styles based on several musical factors such as dynamics, timbre and vibrato [6] is important for music education, music expression analysis and music synthesis. To verify if the selected musical factors or models work effectively to represent a performer’s characteristic style, performer identification becomes an important topic in music signal analysis.

The music informatics literature is rich in works focusing on musical expression analysis and performer identification. Stamatatos and Widmer proposed a set of features like time deviation and melody lead [4] that capture aspects of pianists’ individual style. Ramirez et al. [15] developed a machine learning approach to identify Jazz saxophonists by analyzing the pitch, timing, amplitude and timbre of individual notes. Similarly, Bresin [3] used articulation features to analyse different pianists’ expressions. Particularly, there are prior works on violin expression analysis and violinist classification. Li et al. [8] selected duration, dynamics and vibrato features to classify expressions using Support Vector Machines (SVMs). Molina et al. [11] proposed an approach for identifying violinists in monophonic recordings using a musical trend-based model. Shih et al. [19] used articulation and energy features to compare different playing styles of Heifetz and Oistrakh.

Among the influential factors of music performance, vibrato plays an important role in the performance of singing voice, flute and bowed-string instruments, and is frequently used to enhance selected notes and make them more prominent [12].

Although there is no clear definition of vibrato, it can be broadly described as the musical voice’s “periodic oscillation in pitch” [18]. In violin performance, the characteristics of vibrato mostly depend on the performer’s finger movement on the fingerboard, which is strongly related to the performer’s playing habits. Moreover, vibratos are ubiquitous in violin music, while vibrato features including vibrato rate and extent are seldom influenced by the differences between musical instruments and recording conditions. Thus we choose vibrato features as a characteristic factor to identify different master violinists.

To our knowledge, most previous works attempted violinist identification using features of pitch, timing, energy or vibrato amount in a music piece, while characteristic vibrato features were not considered in this task. Additionally, prior vibrato analysis studies [5], [13], [14], [23] extracted vibrato features from fundamental frequency trajectories of notes in monophonic music segments. In this paper, we propose two approaches that can potentially identify performers in polyphonic music using: *i*) vibrato feature distribution similarity (VF-DS), and *ii*) adaptive wavelet scattering (AWS). The former contains a high-level feature extractor and uses feature distributions to model performers’ individual vibrato playing. The latter, also known as the adaptive time–frequency scattering, is a variant of the scattering transform [10]. It was originally proposed in [22] for representing periodic modulations, including vibratos, in monophonic music. For the polyphonic scenario, we modify the AWS and apply it for the first time to performer identification.

A flowchart of the study is shown in Figure 1. We first construct a dataset from performances of nine master violinists, then extract the predominant melody using the MELODIA algorithm [17]. Two kinds of vibrato features are extracted and performer identification results are obtained using two proposed methods shown by the alternative paths in the diagram. Finally the results are compared.

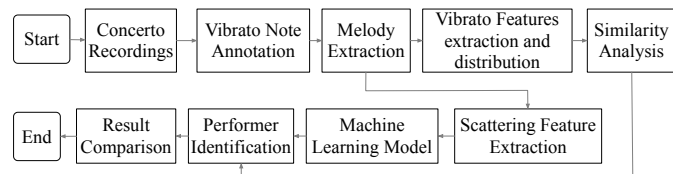


Fig. 1. Flow chart of the proposed systems.

## II. DATASET

In this paper, we use 5 concerto recordings. A concerto is a musical work that focuses on a solo instrument accompanied by an orchestra. The violinist’s personal style is clearly shown while playing a concerto. Compared to solo music (like unaccompanied sonata), orchestral music pieces are more commonly listened to. Every violinist brings an individual style to the performance leading to variations in tempo, intensity and vibrato. For example, Heifetz plays the Beethoven D major violin concerto (III) faster than any other performer which, perhaps by intent, brings an “unemotional and cold” feeling to listeners and differentiates his performances.

We select five concertos written by five well-known composers listed in Table I. 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 violin players. We introduce the dataset annotation process in [24]. In this study, extending previously annotated monophonic vibrato notes, we labelled single vibrato notes with accompaniment to enlarge the dataset. Details of the recordings and the amount of annotated data in each movement are listed in Table I. The total amount of vibrato annotations for each performer is 248.

TABLE I  
CONCERTO VIBRATO NOTE DATASET. WE ANNOTATED THE VIBRATO NOTE SEGMENTS FROM THE ORIGINAL RECORDINGS, ‘# ANNOTATIONS’ REFERS TO THE NUMBER OF VIBRATO NOTE ANNOTATIONS IN EACH MOVEMENT.

Composer	Concerto Name	Movement	# annotations
L. V. Beethoven	Violin Concerto in D major, Op.61	I	21
		II	26
		III	4
J. Brahms	Violin Concerto in D major, Op.77	I	11
		II	6
		III	4
F. Mendelssohn	Violin Concerto in E minor, Op.64	I	13
		II	47
		III	3
P. I. Tchaikovsky	Violin Concerto in D major, Op.35	I	26
		II	7
		III	17
J. Sibelius	Violin Concerto in D minor, Op.47	I	23
		II	24
		III	16

## III. METHODOLOGY

There are two branches in Figure 1 after the “Melody Extraction” step which represent two different vibrato feature extraction methods. These are summarised first while details of each step are given in the following subsections.

### A. Melody Extraction

To capture vibrato features from polyphonic notes, the first step is obtaining the main melody from the audio signal. We extract the predominant melody from annotated recordings using MELODIA [17].

### B. Identification using vibrato feature distribution

Vibrato is typically characterised in terms of two factors: the amount of pitch variation (“vibrato extent”) and the speed with

which the pitch is varied (“vibrato rate”) [20]. To characterise vibrato, we extract four note-level vibrato features: average vibrato extent (AE), average vibrato rate (AR), standard deviation of vibrato extent (SE), and standard deviation of vibrato rate (SR). All features are computed from the extracted melody. The specific feature extraction methods are presented in Section III-B2. In order to describe the characteristic vibrato from different performers, we use a histogram to model such features for each performer, which is shown in Sec III-B3. Finally we compute the feature distributions of each performer, then a violinist classification method based on distribution similarity analysis is provided in Section III-B3.

1) *Data pre-processing*: MELODIA designates segments without main melody as 0Hz. These are left out from our analysis. However, the curve exhibits noise and artefacts near the note peaks and troughs which are not caused by vibrato as shown in Figure 2(a). The signal is therefore smoothed to obtain more reliable vibrato features using a zero-phase Butterworth low-pass filter. This avoids the influence of phase delay. The smoothed signal is shown in Figure 2(b). In case some small fluctuations remain around the boundaries, we address this issue in the following steps.

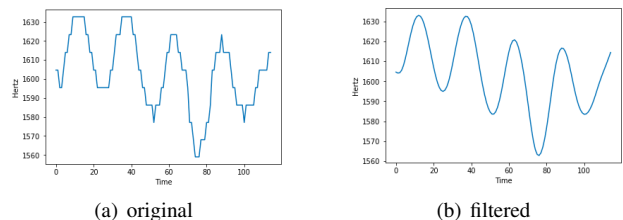


Fig. 2. Vibrato note pitch curve before and after smoothing.

2) *Vibrato feature extraction*: **Vibrato Extent**: In every period of the pitch curve, the instantaneous vibrato extent is considered to be the distance of vertical components between an adjacent peak and trough. The average and standard deviation of the vibrato extent are calculated from all instant vibrato extent values within a note. First we find the location of every peak and trough contained in the pitch curve by locating maxima and minima in the smoothed melody data in each period. We then calculate the absolute frequency distance between successive peaks and troughs to obtain the instantaneous vibrato extent. The collection of note-level instant vibrato extents are used to calculate the AE and SE features for all annotated notes.

**Vibrato Rate**: After obtaining the locations of every peak and trough in the pitch curve of a note, the vibrato rate features can be calculated. We first find the times of peaks and troughs in the pitch curve. The interval between adjacent peaks and troughs is considered a half period  $t_h$ , then the rough instant vibrato rate in every half period is calculated using  $VR = 1/(2 * t_h)$ . The note-level average vibrato rate (VR) is calculated as the mean value of all instant vibrato rates within a note. Despite the pitch curve is smoothed before feature extraction, oscillations or noise that are not caused by vibrato remains a problem. We consider a heuristic to eliminate

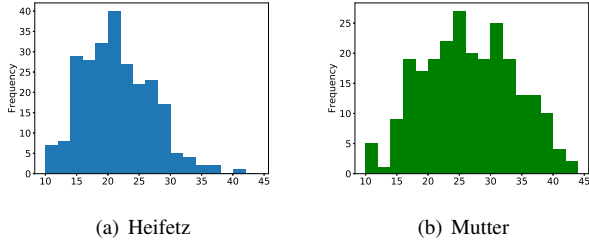


Fig. 3. Distribution of two performers' average vibrato extent (Cents).

the effect of this. In general, the range of the vibrato rate is 2 Hz to 15 Hz, and the range of vibrato extent is between 9 cents and 50 cents, which is also used in [23]. After extracting the rough instant vibrato extent and rate at the note-level, we discard values outside these ranges.

3) *Feature Distribution and similarity calculation.* We use histograms to model feature distributions assuming these provide compact representations of the violinists' style, which we can use later for identification. Figure 3 shows the global distribution of the AE feature for Heifetz and Mutter, while the x axis means the range of vibrato extent(cents), and the y axis means the frequency. It can be seen that Heifetz prefers to use the vibrato in a smaller scale while Mutter's vibrato extents are broader. Based on similar observations across different performers we assume that the feature reflects an important aspect of the vibrato characteristics.

In order to quantify these differences, we calculate the similarity of distributions of each feature for all performers using the Kullback-Leibler (KL) divergence [7], presented as  $D_{KL}(P||Q)$ . This corresponds to the likelihood ratio between two distributions and tells us how well the probability distribution  $Q$  approximates the probability distribution  $P$  by computing the cross-entropy minus the entropy. For classification, the KL divergence is calculated between vibrato 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 Sec. IV-A.

### C. Identification using adaptive wavelet scattering

The dominant band-based AWS in [22] extracts modulation information by adaptively decomposing the signal around the frequency band with maximum acoustic energy. Motivated by the polyphonic nature of violin concertos, we modify the dominant band-based AWS into the *melody-based AWS*, where the latter defines the decomposition trajectory based on the extracted melody in Section III-A.

1) *Theory.* Let  $\psi_\lambda(t)$  denote the wavelet filterbank obtained from a mother wavelet  $\psi(t)$ , where  $t \in \mathbb{R}$  is the time variable and  $\lambda \in \mathbb{R}$  is the log-frequency variable of  $\psi_\lambda(t)$ . Starting from an audio waveform  $\mathbf{x}(t)$ , we convolve it with  $\psi_\lambda(t)$  and take the modulus, and obtain a two-dimensional representation called *scalogram*:

$$\mathbf{X}(t, \lambda) = |\mathbf{x} * \psi_\lambda|(t). \quad (1)$$

Averaging  $\mathbf{X}(t, \lambda)$  by a lowpass filter  $\phi_T(t)$ , we obtain the *first-order scattering transform*:

$$\mathbf{S}_1 \mathbf{x}(t, \lambda) = \left( |\mathbf{x} * \psi_\lambda| * \phi_T \right)(t), \quad (2)$$

which is invariant to time-shifting and time-warping.  $T$  is the averaging scale. Due to the harmonic nature of violin music, one harmonic partial sufficiently captures all the characteristic information of vibratos: rate, extent, and shape. Therefore, we propose to extract vibrato features corresponding only to the frequency bands around the extracted melodies rather than decomposing all frequency bands in the scalogram as the standard scattering transform [10]. This provides a compact representation and is more computationally efficient. We denote the centre frequency of each band in the first-order scattering transform as  $f_\lambda(t)$  and the frequency of the extracted melody as  $f_{\text{melody}}(t)$ . For each time frame, we first localise the melody to the frequency band in the scalogram with the closest frequency value, and obtain the melody trajectory:

$$\lambda_{\text{melody}}(t) = \min_\lambda \left| f_\lambda(t) - f_{\text{melody}}(t) \right|. \quad (3)$$

Vibratos are frequency modulations, which may spread over several frequency bands. To capture the vibrato extent information, we introduce an  $N$ -band tolerance symmetrically centred at the melody trajectory.  $N$  is the total number of frequency bands decomposed. The *melody-based decomposition trajectory* is then expressed as:

$$\Lambda(t) = \left\{ \lambda_{\text{melody}}(t) + n \mid -\frac{N-1}{2} \leq n \leq \frac{N-1}{2} \right\}. \quad (4)$$

Locating this decomposition trajectory to the scalogram, we decompose only a subset of the scalogram:  $\mathbf{X}_\Lambda(t, n) = \mathbf{X}(t, \lambda_{\text{melody}}(t) + n)$ . Convoluting  $\mathbf{X}_\Lambda(t, n)$  with another wavelet filterbank  $\psi_{v_t}(t)$ , taking the modulus and averaging, we obtain the *melody-based adaptive wavelet scattering*:

$$\mathbf{S}_2^{\text{adapt}} \mathbf{x}(t, v_t) = \left( |\mathbf{X}_\Lambda * \psi_{v_t}| * \phi_T \right)(t), \quad (5)$$

where  $v_t \in \mathbb{R}$  is the log-frequency variable of  $\psi_{v_t}(t)$ .

To capture only the temporal variation regardless of the absolute energy of the waveform, we normalise the second-order scattering coefficients over the first-order coefficients and take the logarithm of the normalised coefficients [2], which forms the representation for our performer identification system. We use Morlet wavelets throughout the paper for convolutions. This is because Morlet wavelets have an exactly null average while reaching a quasi-optimal tradeoff in time-frequency localisation [9].

2) *Feature extraction.* Starting from the note-level audio recordings, we extract the characteristic information of vibratos by setting the AWS hyperparameters as follows. Motivated by the vibrato rate of 2-15 Hz in violin music, we set the averaging scale  $T = 2^{15}$  (in samples which corresponds to 743 ms at a sampling rate of 44.1kHz) to capture the rate information. The filters per octave in the first-order scattering  $Q_1 = 16$  are applied to ensure a resolution finer than one

semitone. The extracted melody is then allocated to the nearest frequency band in the first-order scattering transform. Since the extracted melody has higher temporal resolution (3 ms) than that of the scattering transform (186 ms), we use the median per  $186/3 = 62$  melody values. We use  $N = 7$  tolerance bands symmetrically centred at the melody trajectory and filters per octave in the second-order scattering  $Q_2 = 8$  according to preliminary experimental results. The range  $M$  (in Hz) of the modulation rate is useful for reducing the dimension of AWS. Setting  $M$  into an interval, the AWS extracts only the coefficients corresponding to this range. An interval larger than the modulation rate range provides some harmonics in the modulation representation. We use  $M = [0, 50]$ , which results in a feature dimension of 154. The frame size  $h$  (in samples) is inversely log-proportional to an oversampling parameter  $\alpha$ , whereby  $h = T/2^\alpha$ . This is designed to compensate for the low temporal resolution resulting from the large  $T$ . We use  $\alpha = 2$  for all the experiments which corresponds to a frame size of  $h = 186$  ms.

#### IV. EXPERIMENTS AND RESULTS

##### A. Classification using vibrato feature distributions

In order to avoid overlaps between the training and test sets, we use movement-level and concerto-level leave-one-group-out cross validation in classification. In each fold, we designate recordings of one concerto played by all 9 performers as the test set while the remaining recordings are placed into the training set. This eliminates piece overlap between the training and test sets. Each concerto in the test set includes 3 movements. Single labels are assigned in the test set at two levels: concerto-level and movement-level. We then compute the KL divergence between each feature’s distribution from the test performer and the same features for every performer in the training set. Similarity results for vibrato characteristics based on four features are obtained between the test performer and every performer in the training set.

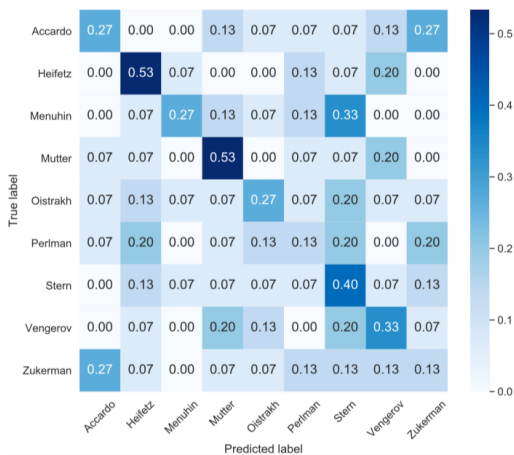


Fig. 4. Normalised confusion matrix for violinist identification using vibrato feature distributions (movement level).

Finally, we compute the mean of normalized KL divergences for four features. We use linear combination with equal

weights to fuse similarity estimates for the distributions of four different vibrato features. During evaluation, leave one group out cross validation with five folds are used to calculate the KL divergence for every group of data. The similarity estimates of feature distributions in every fold are combined for all features using the approach shown in Eq. (6):

$$KL_{overall} = \sum_{n=1}^{|\Theta|} w_n KL_{\Theta_n}, \quad (6)$$

where  $\Theta = \{V_1, V_2, V_3, V_4\}$  with  $V_1, \dots, V_4$  denoting the sets of statistical models corresponding to four kinds of vibrato features (AE,AR,SE,SR) computed separately. These are denoted as ‘combination’ features (VC).

From the cross validation, we obtain the similarity of vibrato features between every two performers in the dataset and performer identification using one feature or the combination feature are obtained. Figure 4 shows the identification results using the VF-DS method and the combination features. Good performances are achieved in identifying Heifetz, Mutter, and Stern, with confusion coefficients of 0.53, 0.53 and 0.4, respectively. However, the identification for Perlman and Zukerman is less reliable. The best performing single feature is the AE with a macro F-score of 0.278 while the worst is the SR. We evaluate the result using F-score metrics [16]. The macro precision, recall, and F-score for all performers at the movement-level are 0.33, 0.32 and 0.31, overall accuracy is 0.32. Corresponding concerto-level result are 0.35, 0.31 and 0.31 with overall accuracy of 0.31.

##### B. Classification using adaptive wavelet scattering

We develop a violinist identification system using the AWS features of vibrato notes calculated from Section III-C as input. SVMs with Gaussian kernels are used as the classifier due to their good generalisability based on a limited amount of training data [1]. The model parameters to be optimised in the training are the error penalty parameter and the width of the Gaussian kernel. We use consistent parameter grids of  $2^{\{3:1:6\}}$  and  $2^{\{-12:1:-7\}}$  for these two hyperparameters, respectively. In the identification process, we conduct both movement-level and concerto-level leave-one-out cross-validation, as in Section IV-A. The best model parameters selected in the training stage are used for testing. Note that the output of SVMs are the violinist identity for each time frame of the vibrato notes. Therefore we assign performer labels according to a majority vote of the frame labels of all vibrato notes in the movement or concerto. The system outputs one performer label per each movement or concerto.

We evaluate performer identification using the same metrics as in Section IV-A. The macro precision, recall and F-score for all 9 performers obtained at the movement-level are 0.38, 0.24, and 0.23 respectively. The overall accuracy is 0.24, lower than that obtained by the VF-DS method. Fig. 5 shows the normalised confusion matrix of the violinist identification results. Notable confusion is observed between Accardo and other performers. Few movements of Perlman’s are identified

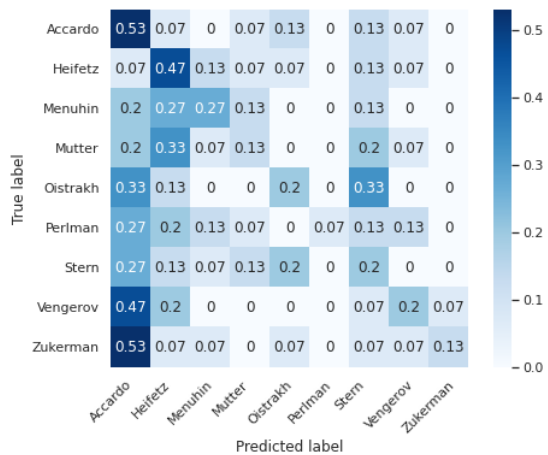


Fig. 5. Normalised confusion matrix for violinist identification using AWS-based vibrato features (movement level).

correctly. The macro precision, recall, and F-score for all 9 performers obtained at the concerto-level are 0.43, 0.29, and 0.27, respectively. The overall accuracy is 0.29.

## V. DISCUSSION AND CONCLUSIONS

From the results obtained in Sections IV-A and IV-B, we find that both the VF-DS method and the AWS method outperform a random baseline, with accuracy improvement of 19.8% and 17.8% respectively on piece-level evaluation. The macro F-score of the former method is better than that of the latter. The VF-DS method is based on note-level vibrato extents and rates, which are high-level and low-dimensional features. Such features can intuitively model individual finger movements on the violin fingerboard when vibratos are performed. Moreover, lower feature dimensionality in VF-DS saves computation cost at the feature extraction and classification stages compared to the AWS method.

Disregarding computational expense, the AWS method is more flexible and promising in three aspects. Firstly, the AWS method does not depend on the number of notes in the test data. The performer can be identified using a majority vote of all frame labels of even a single vibrato note, after the identification machine learning model is trained. In contrast, VF-SD requires a sufficient number of vibrato notes to calculate the feature distribution reliably. Additionally, the AWS method does not require preprocessing the extracted melody (see Section III-B2), which simplifies the system. Finally, with the AWS method, we are able to develop a fully automatic performer identification system. This is because the AWS itself is a vibrato detector [22]. We can first use the AWS transform to detect vibratos in the music piece, and then use the AWS coefficients of the detected vibratos to classify performers. This framework can also generalise to other playing techniques like tremolos, trills, which are also frequently used in violin music and may contain violinists' discriminative information.

In future work, we may fuse features using different weights or design features that correlate with timbre or dynamics. The identification results of both methods depend on melody

extraction accuracy and errors that cannot be corrected in subsequent steps. Using source separation to isolate the violin would potentially improve the result. Classifiers including Decision Trees or Neural Networks may also be tested.

## REFERENCES

- [1] Felix Albu and Dominique Martinez. The application of support vector machines with gaussian kernels for overcoming co-channel interference. In *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No. 98TH8468)*, pages 49–57. IEEE, 1999.
- [2] Joakim Andén and Stéphane Mallat. Deep scattering spectrum. *IEEE Transactions on Signal Processing*, 62(16):4114–4128, 2014.
- [3] Roberto Bresin and Giovanni Umberto Battel. Articulation strategies in expressive piano performance analysis of legato, staccato, and repeated notes in performances of the andante movement of mozart's sonata in g major (k 545). *Journal of New Music Research*, 29(3):211–224, 2000.
- [4] Werner Goebel. Skilled piano performance: Melody lead caused by dynamic differentiation. In *Proc. 6th Int. Conf. on Music Perception and Cognition*, 2000.
- [5] Lynette Johnson-Read, Anthony Chmiel, Emery Schubert, and Joe Wolfe. Performing lieder: expert perspectives and comparison of vibrato and singer's formant with opera singers. *Journal of Voice*, 29(5), 2015.
- [6] Patrik N Juslin and Petri Laukka. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological bulletin*, 129(5):770, 2003.
- [7] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.
- [8] Pei-Ching Li, Li Su, Yi-hsuan Yang, and Alvin WY Su. Analysis of expressive musical terms in violin using score-informed and expression-based audio features. In *ISMIR*, pages 809–815, 2015.
- [9] Stéphane Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, 2008.
- [10] Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.
- [11] Miguel Molina-Solana, Josep Lluís Arcos, and Emilia Gomez. Identifying violin performers by their expressive trends. *Intelligent Data Analysis*, 14(5):555–571, 2010.
- [12] Caroline Palmer and Sean Hutchins. What is musical prosody? *Psychology of learning and motivation*, 46:245–278, 2006.
- [13] Hee-Suk Pang and Doe-Hyun Yoon. Automatic detection of vibrato in monophonic music. *Pattern Recognition*, 38(7):1135–1138, 2005.
- [14] Eric Prame. Measurements of the vibrato rate of ten singers. *The journal of the Acoustical Society of America*, 96(4):1979–1984, 1994.
- [15] 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*, 17(3):356–364, 2007.
- [16] Cornelis Van Rijsbergen. *Information Retrieval (2nd ed.)*. Butterworth-Heinemann, 1979.
- [17] Justin Salamon and Emilia Gómez. Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(6):1759–1770, 2012.
- [18] Carl E Seashore. The natural history of the vibrato. *Proc. of the National Academy of Sciences of the USA*, 17(12):623, 1931.
- [19] Chi-Ching Shih, Pei-Ching Li, Yi-Ju Lin, Yu-Lin Wang, Alvin WY Su, Li Su, and Yi-Hsuan Yang. Analysis and synthesis of the violin playing style of heifetz and oistrakh. In *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, 2017.
- [20] Johan Sundberg. Acoustic and psychoacoustic aspects of vocal vibrato. *Vibrato*, pages 35–62, 1995.
- [21] Sandrine Vieillard, Mathieu Roy, and Isabelle Peretz. Expressiveness in musical emotions. *Psychological research*, 76(5):641–653, 2012.
- [22] Changhong Wang, Emmanouil Benetos, Vincent Lostanlen, and Elaine Chew. Adaptive time-frequency scattering for periodic modulation recognition in music signals. *International Society for Music Information Retrieval Conference*, 2019.
- [23] Luwei Yang, Elaine Chew, and Sayid-Khalid Rajab. Vibrato performance style: A case study comparing erhu and violin. 2013.
- [24] Yudong Zhao, György Fazekas, and Mark Sandler. Identifying master violinists using note-level audio features. In *17th Sound and Music Computing Conference*, 2020.