

Feature-based Characterization of Violin Timbre

Francesco Setragno, Massimiliano Zanoni, Augusto Sarti and Fabio Antonacci

Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB)

Politecnico di Milano

Piazza Leonardi da Vinci 32, 20133 Milano

Email: francesco.setragno@polimi.it

Abstract—Timbral quality of historical violins has been discussed for years. In this paper, we show that it is possible to characterize it from an objective, low-level features perspective. Feature selection algorithms are used to select the features that most characterize historical and contemporary violins. The feature representation of violins is then reduced by means of the T-SNE method. In the low-dimensional space which is obtained, historical violins tend to group together.

I. INTRODUCTION

Violin has been a subject of research for decades. It has been studied from several points of view (acoustic, chemical, structural, etc.). Among them, timbre is certainly one of the most important. Timbre is very hard to define, due to its subjective nature. For this reason, several aspects of the qualities of violins are still to be clarified. Several studies have been proposed [1][2][3], based on the construction of a timbral space where each dimension is correlated with one or more acoustic properties of sound.

Among them, [4] and [5] exploit feature-based analysis for timbre characterization of violins. Low-level features are objective descriptors devoted to capture specific aspects of the sound. Since the timbre is the combination of many factors ranging from acoustics to perception, feature-based analysis resulted particularly suitable for musical instruments characterization [6][7][8]. In [9] the authors take advantage of feature-based analysis for a musical instruments recognition scenario. In their study they propose a method for automatic classification based on a given set of types of instruments: clarinet, cello, guitar, oboe, piano, trumpet, violin. At the best of our knowledge no studies on musical instruments recognition for instruments of the same type have been conducted. In this area, for the violin maker community one interesting aspect in the study of violin sound quality is the timbral characterization of historical instruments and, in particular, the understanding of sound qualities that make historical instruments different from contemporary instruments, if any. The sound of historical violins built by the great masters from Cremona - Stradivari, Guarneri, Amati - are considered as the pinnacle of the violin making art and, after several centuries, they are still used as a model by contemporary violin makers. For this reason they are still under the spot and their sound is subject to many discussions.

In recent studies, through perceptual tests Fritz et al. [10] showed that expert musicians are not always able to distinguish historical from modern violins pointing out the difficulty of

the task. In this paper we study the sound qualities that best allow to discern historical from modern violins through feature-based analysis. We extract a large set of low-level features from a dataset of recordings of historical and contemporary violins. A set of correlation studies is then performed through feature selection algorithms. Since the evolution of the features over time is an important element for what concerns timbre perception, we also take into account different parts of the notes envelope separately for this study.

Through the feature extraction procedure, each instrument can be represented by a point in a high-dimensional space where dimensions of the space are the features. This space is very useful for analysis purposes, but hard to visualize. For this reason, dimensionality reduction methods can be used in order to obtain a low-dimension (2D or 3D) space. The visualization can help to better understand, for violin makers, the sound similarity between instruments. In this study we provide a preliminary analysis that exploits a dimensionality reduction method called t-SNE [11] in order to obtain a low-dimensional space where violins can be visualized.

II. METHODOLOGY

A. Recordings

We record a set of 50 instruments: 13 historical violins from the collection of the Violin Museum in Cremona (Stradivari, Guarneri, Amati), 28 high-quality contemporary violins from the Triennale competition and 9 contemporary violins from the violin making school Istituto Antonio Stradivari. We consider the competition violins and the school violins as separate classes because the construction quality between the two sets of instruments is objectively big. The first set includes some of the best instruments in the world, while the other includes instruments built by students with little experience.

The recordings are performed in a semi-anechoic room, using a measurement microphone always placed in the same position with respect to the instrument. The audio is acquired with a sample rate of 44100Hz. All recordings are performed by the same musician and with the same bow. For each instrument, the musician plays the four open strings (each repeated six times), a sequence of notes on every string, a major scale covering all the strings and 6 pieces of classical music including several styles and techniques. Therefore, each recording results in 15 parts. We refer to them as sessions:

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.

We highlight the fact that, for a given instrument, the timbral content in different sessions can vary considerably. For example, in an excerpt with many of notes played fast, the transients have a different impact than with a single note played slowly. For this reasons, the described sessions are analysed separately.

B. Feature analysis

We extract features presented in [7] and others typically used in the Music Information Retrieval field: the four spectral moments (Centroid, Spread, Skewness, Kurtosis), other spectral indicators (Brightness, Rolloff, Flux, Irregularity, Flatness), features related to the distribution of harmonics (Tristimulus coefficients, Odd-Even ratio), two vectorial features describing the spectrum shape (Mel-Frequency Cepstral Coefficients, MFCC, and Spectral Contrast [12]) and some temporal features (Attack time, Attack slope, RMS energy and Zero-crossing rate). We refer to [13], [14], [15], [16], [17] for a detailed explanation of these features.

The audio files are processed using the following paradigm: each file is divided into short overlapping frames (40 ms each, 50% overlap), and for each frame the low-level features are extracted, resulting in a long feature vector. The Root-mean square energy (RMS) vector is used to select and discard the silence frames, which strongly affect the low-level features value. Points where the RMS crosses a very low threshold τ_1 are selected as the beginning and the end of notes. The samples between notes are discarded.

We decide to take into account both the whole evolution of the note and the steady part only. Indeed, the timbre information contained in the steady part is different from the one contained in the decay or the attack of the sound. For each note, a local threshold τ_2 is determined as the mean of the RMS energy in that region. The steady part of the note is selected as the portion of the note whose RMS is higher than τ_2 . The portion of the note that goes from τ_2 to the silence is the decay part. Figure 1 summarizes this procedure. We decide to analyse both the whole notes and the steady parts only, since we noticed that the decay part has a great impact on the features value for the historical violins.

Once the silence is removed and the notes (or part of them) are selected, the mean value of the features is computed for each session. Therefore, each session results in a matrix $N \times M$, where N is the number of violins and M is the number of features.

C. Feature selection

In order to discover the features that best characterize the timbre of historical and contemporary violins, we run five different feature selection algorithms. These algorithms select features based on a classification task, where the classes in this case are historical, modern and school violins.

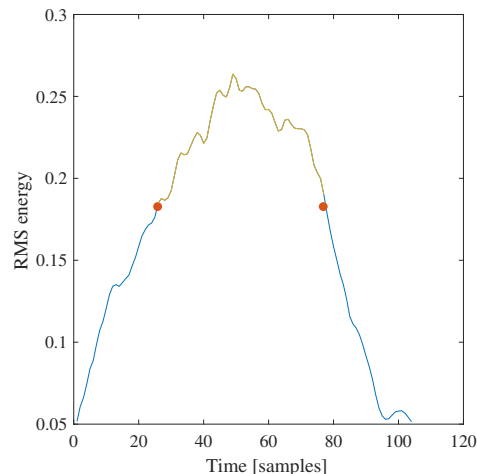


Fig. 1. In this Figure the RMS energy related to the execution of one note is represented. The red dots indicate where the energy crosses the local threshold. The yellow portion of the plot represents the steady part of the sound.

The first three algorithms are provided by Python's *sklearn* toolbox [18]. SelectKBest and SelectPercentile select the K features and a given percentage of the features with the highest score, respectively, according to a statistical test (ANOVA). We try different values of K and different percentages. False Positive Rate (FPR) selects the p -values below a threshold α based on a FPR test. We also use two methods that provide feature ranking and assign a score to each feature: one based on a Forest-Of-Trees, illustrated in [19], and one called Relief [20].

The outputs of these algorithms are compared. Since the timbral property of an instrument is dependent on what is played, we make the comparison separately for each session.

D. Dimensionality reduction

For visualization purposes, the feature vectors can be reduced to a low dimensionality. Methods such as Singular Value Decomposition (SVD) or Principal Component Analysis (PCA) are able to project high-dimensionality vectors into a lower dimensionality space. In this study we use the t -distributed Stochastic Neighbor Embedding (t -SNE) method, illustrated in [11]. This method is used in a wide range of fields and is well-suited for visualizing high-dimensionality data. In our case, the output of the t -SNE algorithms is a 3D vector representing the projection of the features into a 3D space. Violins from our dataset are then plotted in such space. This could be useful to intuitively compare a specific violin to a set of instruments from a low-level point of view.

III. RESULTS

In this section we illustrate the results we achieved.

A. Feature selection results

The feature selection algorithms have been run for every session. Figure 2 shows how many of them, among the five used, selected a given feature. It can be seen that some

coefficients of the MFCC and the Spectral Contrast have a great impact, especially for what concerns the scale and the musical excerpts. As far as regarding the open strings and the single notes, where the steady part of the sound is predominant, the features related to the distribution of frequencies (Centroid, Skewness, Rolloff, Kurtosis, Flatness) are the most important. The Spectral Irregularity, related to the variations in the spectrum, appeared to be important as well.

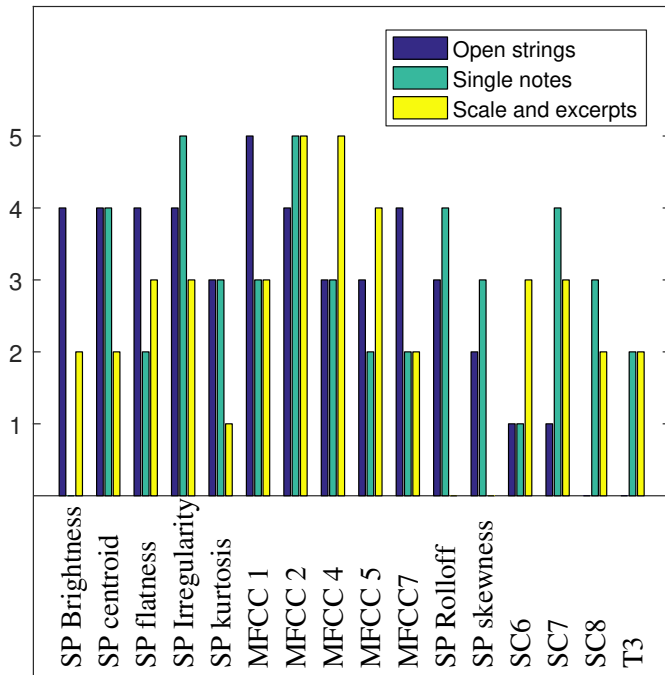


Fig. 2. Number of algorithms that chose each feature for different sessions. SC stands for Spectral Contrast, while T3 stands for Tristimulus (first coefficient)

By examining the spectrograms related to open strings, an important aspect related to the evolution of the notes emerges. In the decay part of the sound (i.e. the period of time in which the energy goes to the steady value to zero), only the fundamental frequency and the low harmonics remain. For what concerns historical violins and some contemporary ones, these harmonics retain a big amount of energy in this phase. When computing the mean value of the low-level features across the whole note duration, this strongly affects the result. For example, the Spectral Centroid results in a very low value in the decay part of the note with respect to the steady part. Figure 3 shows an example of this phenomenon, with two spectrograms related to a historical violin and a school violin, respectively. It can be noticed that, in the decay phase (after the detachment of the bow), the power of the fundamental and the first harmonics remains high for a few seconds for the historical violin (losing about 3 dB with respect to the note attack), while it highly decreases for the contemporary one (losing more than 30 dB).

In order to take this phenomenon into account, we ran the feature selection algorithm by considering only the steady part of the sound. As it is possible to see in Figure 4,

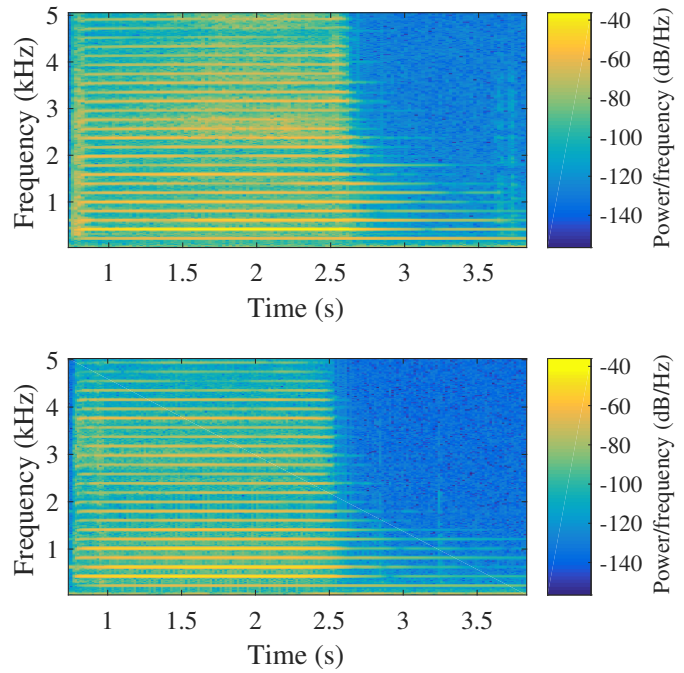


Fig. 3. Comparison of two spectrograms related to the execution of a note on the G string with an historical violin (top) and a school one (bottom). It is possible to notice the steady part and the decay part of the sound, where only the low harmonics remain.

there is less agreement between different algorithms for what concerns the excerpts and the single notes. Also in this case, among the most important features we find some MFCC and Spectral Contrast coefficients. The features related to the frequency distribution (Rolloff, Centroid, Brightness) are no longer selected. This means that, for these features the main differences between historical and contemporary violins lie in the decay of the sound. Moreover, the Spectral Flux, related to the variability of the spectral components over time, is selected.

The features related to the attack of the sound did not appear to be relevant for this characterization task. Therefore, results related to the attack phase are not shown in this paper.

B. Validation with a classification task

In order to show that the selected features are relevant in the discernment of historical and contemporary instruments, we ran a classification task and we examined the performance of the classifier using both the whole set of features and the selected features. The input of the classifier is the vector of features, extracted with the procedure explained in the previous section. Here, we consider the whole note envelope. The output is the class of the sample (Historical or Contemporary). We used the Support Vector Classifier [21] for this task. The dataset was split as follows: 70% for the training set and 30% for the test set. The error was computed as the percentage of misclassified samples. Results are shown in Table I. Feature selection improves the performance of the classifier, especially when the open strings are considered.

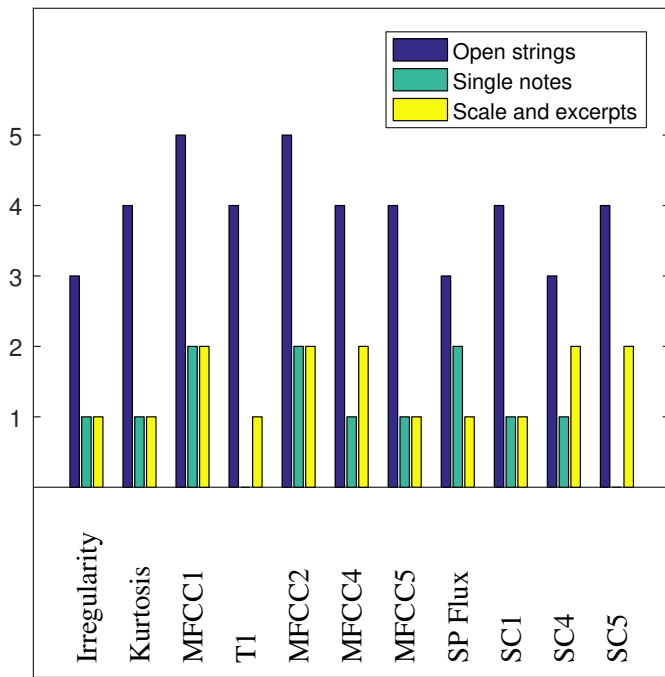


Fig. 4. Number of algorithms that chose each feature for different sessions, considering only the steady part of the sound.

In this study we did not consider a possible impact of the played pitch. In sessions different from the open strings, where different pitches are present, the effect of the feature selection is less clear. Nevertheless, the classification results are good, meaning that low-level features allow to discern historical from contemporary instruments.

C. Validation with visualization

The t-SNE method was used to reduce the dimensionality of the feature vectors containing the selected features. The space was reduced to a dimensionality of 3. In Figure 5 results are displayed for the open strings. It is possible to see that the selected features allow to discern the historical violins. In particular, the separation between historical violins and school violins is clear (especially for G string and D string).

The same result is not achieved with musical excerpts. This can be due to the fact that the variability of the low-level features during the execution (different pitches are played) is more significant than the difference between different instruments.

D. Analysis of the steady part

From the previous results it is clear that the major difference between historical violins and (most) contemporary ones lies in the decay phase of the notes. We examined the values of the low-level features in the steady part of the notes in order to check if there are some features that present remarkable differences. For what concerns the G and A strings, the value of the first and second Tristimulus coefficient appears to have a distribution that varies from historical to contemporary instrument. In Figure 6, this distribution is depicted.

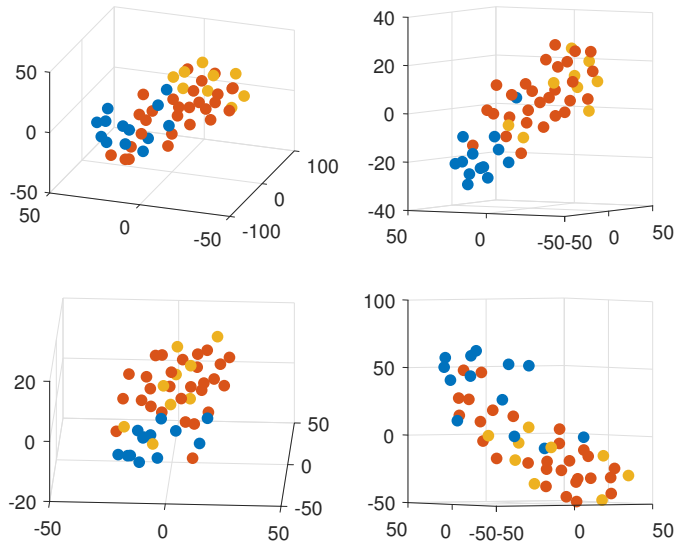


Fig. 5. 3D representation of the feature space obtained using t-SNE on the selected features (clockwise from top-left: G string, D string, E string, A string). Blue dots represent historical violins, red dots represent contemporary good violins and yellow dots represent school violins.

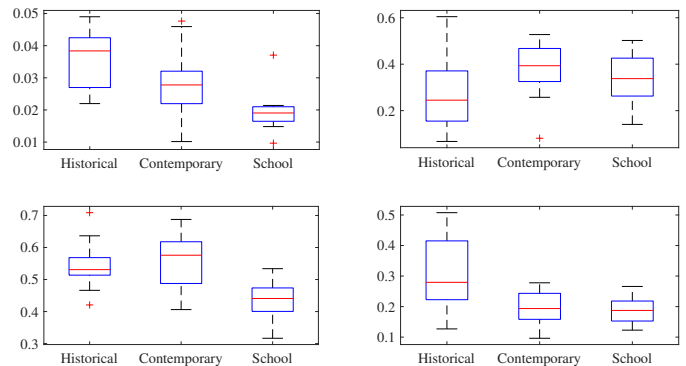


Fig. 6. Values of the Tristimulus coefficients 1 (top) and 2 (bottom), both for G string (left) and A string (right), using the steady part of the notes.

Figure 7 shows the distribution of the Spectral Flux for the high strings (A and E). In this case, good violins (historical in particular) tend to have a higher Spectral Flux, meaning that they exhibit a quicker variability in the spectrum during the execution of a note.

IV. CONCLUSIONS

In this study we showed that historical violins exhibit some low-level objective properties that allow to distinguish them from modern instruments.

50 violins were recorded and several low-level features were extracted. By means of five feature selection algorithms, the most characterizing features were chosen. A dimensionality reduction technique was employed to build a 3D visualization space where recorded violins could be arranged. Results show that, at least for steady sounds where transients do not have a big impact, it is possible to distinguish historical violins from modern ones. In particular, the decay phase of the

Session	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
AF (%)	6	3	4	4	13	17	17	18	11	12	6	13	13	6	21
FS (%)	2	1	3	2	2	15	18	16	10	11	10	10	18	3	19

TABLE I

CLASSIFICATION ERROR WITH THE SUPPORT VECTOR CLASSIFIER, BOTH WITH ALL FEATURES (AF) AND WITH FEATURE SELECTION (FS)

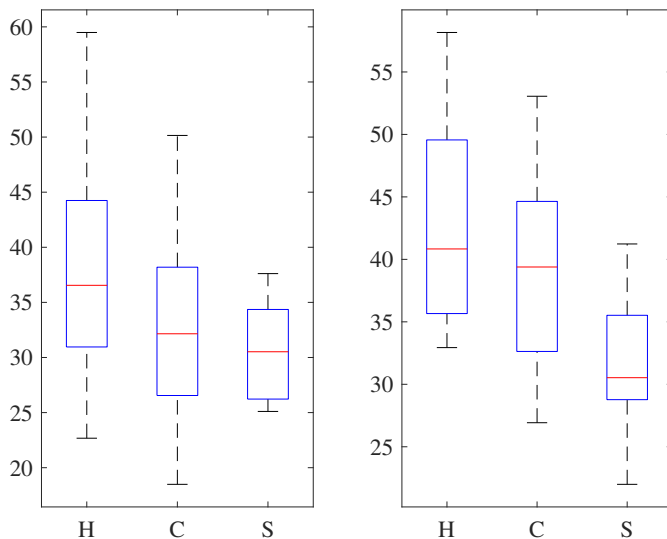


Fig. 7. Values of the Spectral Flux, both for A string (left) and E string (right), using the steady part of the notes.

sound appeared to have a great role in characterizing historical violins, which retain more energy in the low harmonics than the contemporary ones.

Future studies will focus on how the timbral differences between violins change at different pitches, i.e. if there is a dependency between timbre and pitch.

ACKNOWLEDGMENT

This research activity has been partially funded by the Cultural District of the province of Cremona, a Fondazione CARIPLO project, and by the Arvedi- Buschini Foundation. The authors are also grateful to the Violin Museum Foundation, Cremona, for supporting the activities of timbral acquisitions on historical violins of its collection.

REFERENCES

- [1] A. C. Disley, D. M. Howard, and A. D. Hunt, "Timbral description of musical instruments," in *9th International Conference of Music Perception and Cognition*, 2006.
- [2] J.M.Grey, "Multidimensional perceptual scaling of musical timbres," *Journal of the Acoustical Society of America*, 1977.
- [3] S. McAdams, S. Winsberg, S. Donnadiou, G. D. Soete, and J. Krimphoff, "Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes," *Psychological Research*, 1995.
- [4] J. A. Charles, D. Fitzgerald, and E. Coyleo, "Violin Timbre Space Features," in *Irish Signals and Systems Conference, 2006. IET*, 2006, pp. 471–476.
- [5] E. Lukasik, "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] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, "The timbre toolbox: Extracting audio descriptors from musical signals," *The Journal of the Acoustical Society of America*, vol. 130, no. 5, pp. 2902–2916, 2011.
- [8] A. Eronen and A. Klapuri, "Musical instrument recognition using cepstral coefficients and temporal features," in *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on*, vol. 2. IEEE, 2000, pp. II753–II756.
- [9] 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.
- [10] 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.
- [11] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. Nov, pp. 2579–2605, 2008.
- [12] 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*.
- [13] T. S. H.G. Kim, N. Moreau, *MPEG-7 Audio and Beyond. Audio Content Indexing and Retrieval*. John Wiley & Sons Ltd, 2005.
- [14] 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.
- [15] K. Jensen, "Timbre models of musical sounds, tech. rep. rapport 99/7," University of Copenhagen, Tech. Rep., 1999.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] P. Geurts, D. Ernst, , and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63(1), pp. 3–42, 2006.
- [20] K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm," in *AAAI*, vol. 2, 1992, pp. 129–134.
- [21] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.