Classification of Violin or Viola repertoires using Frequency Histogram and Mel-Frequency Cepstral Coefficients (MFCC)

Abstract

Due to the monumental growth in the field of machine learning, a lot of research work has been done to retrieve musical information for analytics and application purposes. The subfield, musical instrument classification is the main focus of the upcoming research project, requiring relevant features to be extracted. The two instruments will be focused are violin and viola. Since the two instruments belong to the same string family, it is hard for humans and even experts to distinguish them due to the similarities of their innate timbral attributes. In this research, the two features, namely frequency histogram and Mel-frequency Cepstral Coefficients (MFCC) will be capitalized on to perform the classification task. In the meanwhile, the relative strengths and shortcomings of each feature are being compared.

Index Terms - Machine learning, Musical Information Retrieval, Musical Instrument Classification

1 Introduction

With the rapid advancement in the field of machine learning in recent years, music information retrieval and analysis have gained great interest and attention from researchers. Splendid music information retrieval (MIR) systems will possess salient social and commercial impacts. There are many applications which derived from MIR such as automatic audio searching, video-scene annotation and analysis , songs playlist generation [1], to name a few. To be more specific, musical instrument classification applications are able to analyse music songs or passages to distinguish the music genre or music era of which the specific song is composed.

The data mining and audio signal processing techniques have been capitalized on for the research on music genre classification [2], content-based music retrieval [3], as well as musical instrument classification [4], the sole focus of this work. The general idea of musical instrument classification is to automatically recognize the instruments playing music.

Over the past few years, the research in the field of audio signal processing has prioritised speech recognition over musical instrument classification, not least in the endeavour to improve mobile applications via greater user experience. However, both problems require correct feature extraction techniques and follow the same multi-class classification scheme. Musical instrument classification is not an easy task as there are relatively few features in nature that can be capitalized on to perform the corresponding classification. The task is considered harder, especially for the case of using data sets consisting of polyphonic sounds where several instruments playing music simultaneously. This work will focus on Violin and Viola performing repertoires consisting of both monophonic and polyphonic sounds.

In this work, two main feature extraction techniques have been proposed to select the relevant features for the descriptors. Then, supervised machine learning classification algorithm, random forest ensemble method is used to perform classification. By comparing the results of two features used , the characteristics and distribution of the features can be further explored. The most frequently used notes of violin repertoire tend to be higher than the viola repertoire which are caused by the strings and playable notes of the instruments. However, most researchers did not take into account the structure of the whole song into account as the compositions for violin and viola of different musical eras tend to be different. Moreover, the range of notes in these repertoire become wider from era to era as composers have a more open mindset and become more creative over time. So, repertoires are broken into different eras to be evaluated.

The **identified gap** in the existing literature is most works do not take the musical era and structures of repertoires into account as well as the exclusion of frequency histogram as attribute used. Thereby, with the **objective** of creating a model to classify violin and viola, this research **proposes** to use an unprecedented combination of frequency histogram and MFCC features as the main attributes. Due to the physical design, Violin can play some higher notes which cannot be played on Viola. Based on this setting, the Violin repertoires will most probably has higher notes than Viola repertoires, making the adoption of frequency histogram a decent feature selection choice.

The performing repertoires ranging from Baroque era to Romantic era will be used as the dataset. Classification will be done on repertoires of different eras separately and collectively so as to identify the importance of compositional structure of repertoires towards the classification. Via validating the **hypothesis** of this research made, we can realize the importance of **frequency histogram** towards classification task. A decent model to classify Violin and Viola can help in nurturing musicians to some extent, improving their musicality as well as music sense. Moreover, identification of violin and viola can further aid music genre classification, repertoire analysis, and even music compositions, contributing to the music industry to some extent.

2 Literature Review

2.1 Related works

Many methodologies have been proposed to perform musical instrument classification by using a myriad of feature extraction schemes as well as classifiers. The naive methods used domain knowledge of audio signal to extract features and performed the following classification using basic algorithms such as support vector machines (SVM). In fact, MFCC features are quite popular in music genre classification, musical instrument classification, as well as speech processing.

For instance, Diment *et al.* [5] used a combination of phase related and MFCC features to perform musical instrument classification using Gaussian mixture model . Their results were good on RWC datasets consisting of solo recordings of instruments, achieving **70.7**%, **84.9**%, and **96**% in classifying 22, 9, and 4 instruments. Eronen [6] evaluated the significance of spectral features, MFCC, spectral centroids and amplitude envelope towards musical instrument classification. Decorrelation on the features was performed using Karhunen-Loeve transform. k-nearest neighbor (k-NN) was used as the classifier and the results were further assessed using cross validation. Eventually, MFCC features had been proved to be the important features. Moreno [7] used a support vector machine to perform classification on 8 instruments and he was able to achieve **70**% accuracy.

Xiong *et al.* [8] performed comparison between MPEG-7 and MFCC features for sports-audio classification, using classifiers such as Gaussian mixture models, AdaBoost, hidden Markov Models (HMMs), and so on. Kaminskyj *et al.* [9] used k-NN classification algorithm on monophonic instrument-sound classification, using features such as constant Q-transform spectrum frequency, MFCC features, spectral centroid, root-mean-square (rms) amplitude envelope, as well as multidimensional-scaling(MDS) analysis trajectories. Then, dimensionality-reduction technique Principal Component Analysis(PCA) was employed, eventually gaining **97**% accuracy in instrument-family classification as well as **93**% accuracy in instrument classification.

In recent years, deep learning models have become popular in many research fields as they can directly extract features themselves and perform classification in one shot. In this case, Li *et al* [10] – passed in raw audio waveforms to a CNN) and achieved **72**% F-micro score on clip-level instrument classification on MedleyDB dataset, whereas the combination of random forest and MFCC features achieved merely **64**% F-micro score.

There is no proper benchmark for musical instrument classification as different work uses different dataset, making the comparison in terms of performance among several approaches very hard. Some researchers collect their own dataset while some researchers use dataset available online such as McGill University Master Samples, to name a few.

2.2 Comparison of Violin and Viola

2.2.1 Body of violin and viola

Figure 1: Physical comparison of viola and violin

From the physical aspect, Violin is smaller than Viola(see in Fig. 1). The violin's body is 14 inches, while the viola's body ranges from 15 to 18 inches, with 16.5 inches being the most popular. When violin and viola players play with bows, vibrations from the strings are transmitted to the top plate and bottom plate through the bridge, and this reverberates within the hollow body, producing the rich, brilliant tone characteristic of the violin and viola. A viola's sound is typically deemed as deep and mellow, and it generally has a slower sound than the violin due to its thicker strings. As viola is a larger instrument with thicker strings and larger sound box, it requires a heavier bow with a firmer technique to produce its rich sounds.

2.2.2 Role of violin and viola in various settings

In an orchestra or string quartet setting, there are two sections of violins, namely the first and second violins, yet there is only one viola section. The violins normally carry the melody, especially the first violins, typically play a more difficult part using higher positions. While the second violins will often support the first violins' harmony by playing it in a lower pitch. They may also frame the first violins by playing a countermelody, or providing a rhythmic or harmonic support to the first violins' melody. On the other hand, viola are perfect foils to play rhythmic or harmonic accompaniments to the violins, leaving the second violins free to contribute to the melody.

Due to the rich and distinctive sound, the violin are viola are an excellent choice for playing most classical solo music. The tones and timbre of the violin and viola make them available to play single notes in their solo pieces without making the melodies played dull. During the Baroque and Classical eras, the word *solo* was arbitrarily equivalent to sonata, and could refer either to a piece for one melody instrument with (continuo) accompaniment, or to a sonata for an unaccompanied melody instrument. Accompaniment is the musical part which provides the rhythmic and harmonic support for the melody. As piano has a wide pitch range and two or more independent musical lines at the same time, it is commonly adopted as accompaniment to violin and viola to produce a rich context to embellish violin and viola.

2.2.3 Ranges of violin and viola

Table 1: Range of notes of instruments

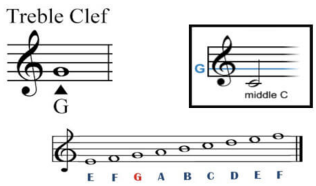
INSTRUMENT	RANGE (MUSIC NOTE)
Viola	C3 - E6
Violin	G3 - A7

The violin strings are G3, D4, A4, and E4, while the viola strings are C3, G3, D4, and A4. Violin is tuned one fifth higher than the viola. The strings of both instruments are tuned one fourth down from each other. Range of violin starts from G3 (196 Hz) to A7 (1760 Hz) while the range of viola starts from C3 (130.81 Hz) to E6 (1319 Hz).

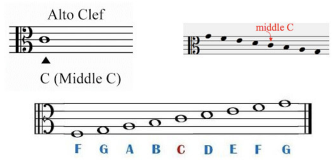
The viola has a mellower and deeper sound as compared to the violin. There are a number of notes that are shared by both instruments, such as the notes on the G3, D4, and A4 strings. However, when these notes are played on a viola, the sound is different as viola carries a more somber tone due to its larger size and its strings are thicker.

Figure 2: Excerpt from Beethoven Symphony No 5 First Movement

As shown by the Fig. 2, the notes played by first violin in an orchestra setting are normally highest, followed by second violin, while the notes played by viola are lowest.



(a) Notes on the Treble Clef



(b) Notes on the Alto Clef

Figure 3: Treble Clef and Alto Clef

2.2.4 Samples of solo pieces

Violin is recognized as the “soprano” voice of any ensemble or orchestra. It is normally played in the Treble Clef. While viola is a middle range alto voiced instrument and it is played in the Alto clef. As the Fig. 3 shown, the notes played by violin is normally higher than viola by using the middle C (C4) as benchmark.

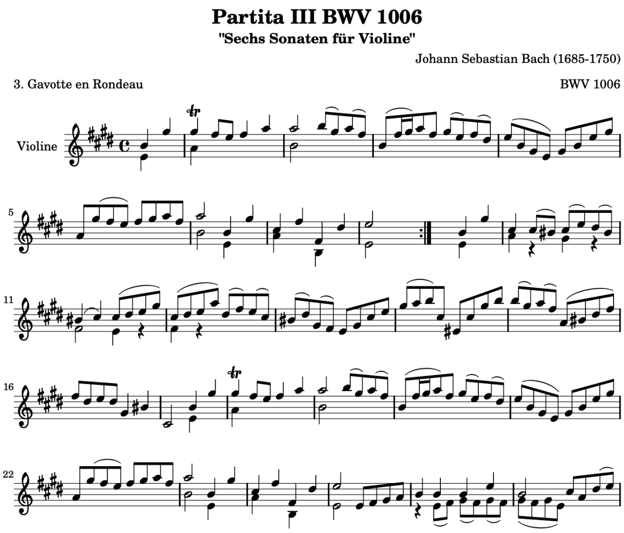


Figure 4: Excerpt from Bach Partita No 3 - 3rd Movement Gavotte en Rondeau

Fig. 4 shows the excerpt from Bach Partita No 3 - 3rd Movement Gavotte en Rondeau and the normal ranges of notes played by a violin in a solo music piece while Fig. 5 shows the detected highest note and lowest note in the aforementioned repertoire, namely D6 and B3.

(a) Highest note detected - D6

(b) Lowest note detected - B3

Figure 5: Highest and lowest notes detected from Gavotte en Rondeau

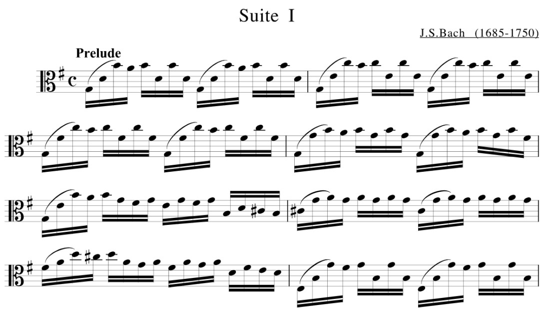


Figure 6: Excerpt from Suite 1 of Bach's 6 Suites for Viola

Fig. 6 shows the excerpt from a Viola Suite and the normal ranges of notes played by a viola in a solo repertoire while Fig. 7 shows the detected highest note and lowest note in the aforementioned viola repertoire, namely G5 and C3.

From the results shown, we can notice the range which is the difference of the highest note and the lowest note detected in both violin and viola music piece is almost the same on mel-frequency scale. However, the highest note and lowest note detected in a violin piece are higher than those detected in a viola piece.

2.3 Chronology of music era

Many people have a misconception that the eras before 20th century are classified as a sole Classical era. However, as Fig. 8 shows, there are indeed 3 main musical eras, namely Baroque era, Classical era, and Romantic era before the 20th century.

2.3.1 Baroque music era

The music of the Baroque [11] music period showed the reflection of the decorative art in the use of ornamentation to embroider melodies. Thick and complex polyphonic texture prevails in many composers' compositions. Vibrant rhythms and expressive dissonances heighten tension in many Baroque works.

Counterpoint is a style of musical writing which involves intertwining

(a) Highest note detected - G5

(b) Lowest note detected - C3

Figure 7: Highest and lowest notes detected from Suite 1

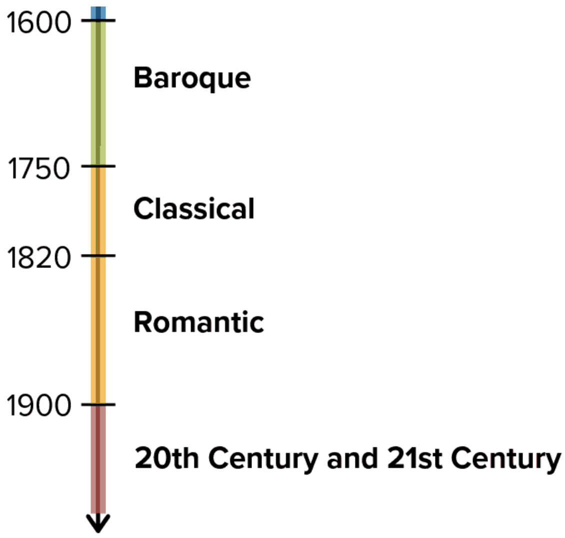


Figure 8: Evolution of music eras

two or more melodies. Each individual melody (also called a ”part”) is an independent, musical idea, and the emphasis is on the separate strands of melody, on the horizontal axis, rather than on the harmony, or vertical axis. This style of writing originated in the Renaissance era, and reached its apotheosis during **Baroque** times.

At the time of Palestrina (Renaissance era), music was written in modes - scales based on different patterns of tones and semitones than the major and minor scales we use today (with some overlaps). The modal system was gradually surpassed by the diatonic (major/minor) system, and by Baroque times most music was diatonic. This happened partly due to developments in tuning systems, which allowed keyboard instruments to be ”in tune” regardless of which key the music was in, thanks to a system called ”equal temperament”.

Table 2: Characteristics of Baroque music

Characteristic	Description
MELODY	A single melodic idea
RHYTHM	Continuous rhythmic drive
TEXTURE	Balance of Homophonic(melody with chordal harmony) and polyphonic textures
TIMBRE	Orchestral strings, winds and harpsichord with very little percussion
DYNAMICS	Abrupt shifts from loud to soft - achieved by adding or subtracting instruments

2.3.2 Classical [12] music era

Trends began to change, and the perpetual motion style gave rise to a style based on short and tuneful phrases. Composers like Mozart, Beethoven, Schubert and others took elements from counterpoint, and used it in a different way to Bach - counterpoint became just one aspect of texture in a work, and it was much less likely to be used in a work as a whole. Contrapuntal sections were often contrasted with homophony (melody plus accompaniment). Mozart and Beethoven chose to employ more contrapuntal textures in their later music, than in their more youthful compositions.

During this era, a new style, highly refined, simple in melodic line and harmonic texture and unified by symmetrical form. Developing during the early Classical period were expanded instrumental forms such as the sonata, allegro and rondo forms. The binary dance movements of the Baroque gave way to the ternary first movements of most Classical period works (sonata, concerto, chamber music, symphony) which comprised three parts: exposition (A) development (B) and recapitulation (A).

Table 3: Characteristics of Classical music

Characteristic	Description
MELODY	Short and clearly defined musical phrases with two or more contrasting themes
RHYTHM	Very defined and regular
TEXTURE	Mostly Homophonic
TIMBRE	four sections in orchestra- strings, woodwind, brass and percussion

2.3.3 Romantic music era

Romantic music [12] focused on subjective emotion and personal experience, national pride, musical richness as well as the flamboyance requiring virtuosic skill. Before this, music was normally stiff and rigid - beautiful, but focused to achieve an almost academic perfection. The rules were all strict. Romantic composers started to bend these rules, playing with new sounds, ideas, and even instruments.

The music of the Romantic era mostly contained warm, personal melodies; expressive indications (espressivo, dolce, con amore, con fuoco,) implied interpretive freedom(rubato) and harmonic colour (new chords such as the ninth) Colour was intensified by improvements in instruments, especially the piano. Performers carried the new music to great heights with the new improved versions of their instruments. During this period exaggerated emotional response was displayed.

Table 4: Characteristics of Romantic music

Characteristic	Description
MELODY	Long, lyrical and wide melodies with irregular phrases
RHYTHM	Frequent changes in both tempo and time signatures
TEXTURE	Almost entirely homophonic
TIMBRE	A variety of tone colour; rich orchestration.

3 Methodology

3.1 Dataset

The instruments included in this work are violin, and viola. The dataset of MP3 data type collected from Youtube consists of repertoires from Baroque, Classical, and Romantic eras. The number of solo repertoires of the Baroque era found are more than the repertoires with piano accompaniment of the Baroque era. While for the Classical and Romantic era, solo repertoires of Violin and Viola are relatively fewer as compared to the Baroque era due to the reason that ensemble, string quartet and orchestra composition were more popular during these eras. In order to prevent imbalance of the dataset for different eras, for each instrument, 10 solo repertoires and 10 repertoires with piano accompaniment were included for each era. Likewise, there is no any repeated repertoire in the dataset other than those transcribed repertoires for Viola. Table 5 shows the hierarchical structure of the dataset.

On the other hand, some compositions are initially made for merely violin. Due to the rise of popularity of Viola and the constraint of the playable notes on Viola, transcriptions were made for some repertoires. For instance, Bach's Partita No 3 Prelude was initially in E major key signature, it was transposed into a version with A major key signature to be played on a Viola.

As some performing repertoires were recorded during a live performance, there were some redundant parts such as speaking, and audience clapping. Thereby, optimal trimmings are done for those repertoires to facilitate the later features extracting parts. The amplitude of piano sounds in non-solo pieces was minimized using Riffstation [13] so that the sounds of Violin and Viola played can be made clearer.

Moreover, most famous repertoires were played by different players ranging from amateur to professional. Thereby, the best version of a specific repertoire was chosen qualitatively by taking into account the recording quality, the amount of noise in the background, the amount of notes which are off-pitch as well as the calibre of the performers.

Table 5: Hierarchical Structure of the Dataset

Era	Solo pieces				Pieces with accompaniment			
	Violin		Viola		Violin		Viola	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Baroque	7	3	7	3	7	3	7	3
Classical	7	3	7	3	7	3	7	3
Romantic	7	3	7	3	7	3	7	3

3.2 Extracting and selecting features

As the timbre of instruments exhibits colorful spectrum innate, several feature extraction schemes have been implemented via the research on time domain and frequency domain spectrum. In this work, the main aim is to find a combination of features which will aid the performance of musical instrument classification tasks at expense of including some redundant features which require more computational resources. The two main features being used are representation of frequency histogram, 13 Mel-frequency cepstral coefficients (MFCC) features. As the timbre of instruments exhibits colorful spectrum innate, several feature extraction schemes have been implemented via the research on time domain and frequency domain spectrum. In this work, the main aim is to find a combination of features which will aid the performance of musical instrument classification tasks. The two main features being used are representation of frequency histogram, and 13 Mel-frequency cepstral coefficients (MFCC) features.

3.2.1 Frequency detected

Table 6: Range of notes of instruments

INSTRUMENT	PLAYABLE RANGE (MUSIC NOTE)
Viola	C3 - E6
Violin	G3 - A7

As shown via Table 6, different instruments have different performing ranges. In fact, this leads to a faster and alternate way of classify musical instruments at first sight. For instance, if a C3 note is played, we can immediately classify it as viola because this note is merely playable on viola. So, the other features might not be needed in this case. However, there are still many overlapping music notes between these 2 musical instruments. For example, a C4 note is playable on the aforementioned instruments. So, the frequencies detected from audio dataset are capitalized on for the upcoming by representing them via histogram visualizations. The occurrences of each note are counted based on the notes found on each audio sample, instead of the counting the notes played throughout the whole music piece. For instance, if a single C4 note (monophonic sound) is played in audio file, the longer the notes being played, the more the occurrence of the corresponding note being counted instead of merely one. The flow diagram of adopting frequency diagram can be found in Fig. 9. The features fed into the classifier are represented via an array consisting of the maximum, minimum, 25th percentile, 75th percentile, mean, mode, and median of frequency(log-based) detected from the repertoires.

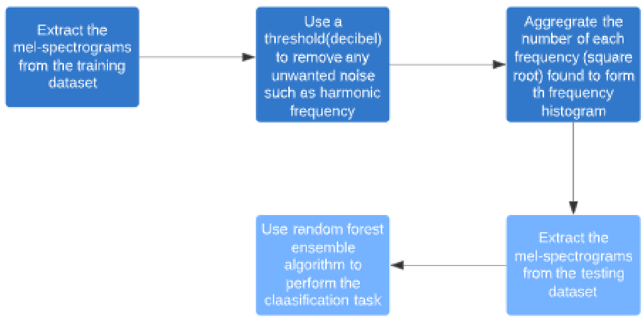


Figure 9: Flow diagram of using Frequency histogram

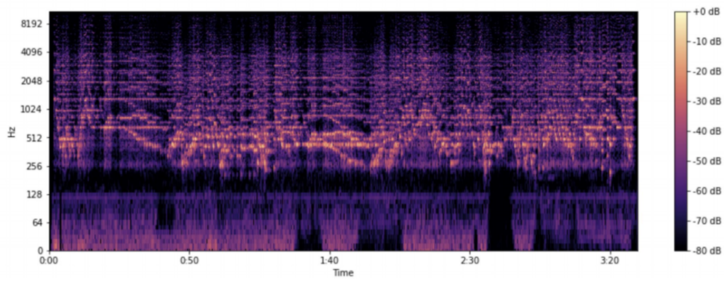


Figure 10: Mel-spectrogram of Partita No 3 Prelude in E Major for violin

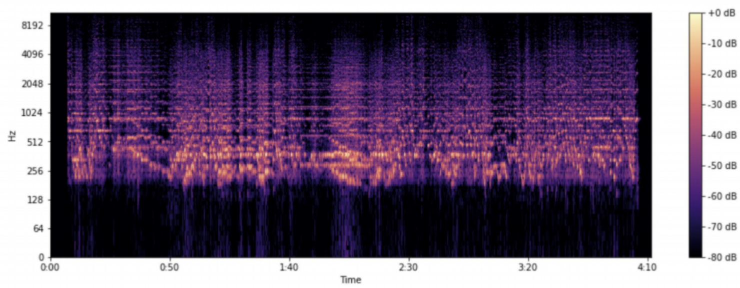


Figure 11: Mel-spectrogram of Partita No 3 Prelude in A Major for viola

Mel spectrogram, the non linear transformation of the frequency scale is prioritized over normal spectrogram so that sounds will be of equal distance from each other on the Mel Scale. For instance, the difference between 200 Hz and 300 Hz is more distinguished than the difference between 4000 Hz and 4100 Hz for the human ear.

As the Fig. 10 and 11 shown, harmonic sound, or more specifically, higher octave notes will be generated when instruments played a fundamental frequency of a certain note. Dominant notes found in violin repertoires are normally higher than those in viola repertoires. Low pass filter needs to be included to attenuate frequencies which are higher than the ground truth frequency played so that the dominant frequencies can be visualized via histograms. Normalization will be done on the audio amplitude by setting the largest amplitude of each audio file as 0 decibel and -10 decibel is used as the threshold to classify the sample note as the distinguished note.

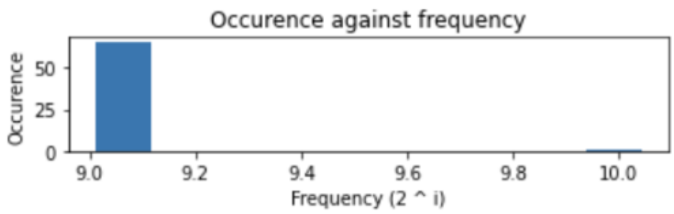


Figure 12: Histogram of frequencies detected for C4 note after preprocessing

Then, the logarithmic value of the remaining frequencies will be used to plot the histogram as shown by Fig. 12. This process is somehow deemed as normalization for audio signals on the basis of the mel spectrogram.

3.2.2 MFCC

In order to derive the spectrogram which is the frequencies detected from audio signals as time changes, Short-time Fourier transform (STFT) needs to be implemented in advance so as to get the phase content and sinusoidal frequency of audio signal as a function of time. The step for the computation of STFTs is to separate the whole audio into smaller segments of equal time length and then the corresponding Fourier transform on each individual segment is computed. The flow diagram of adopting frequency diagram can be found in Fig. 13.

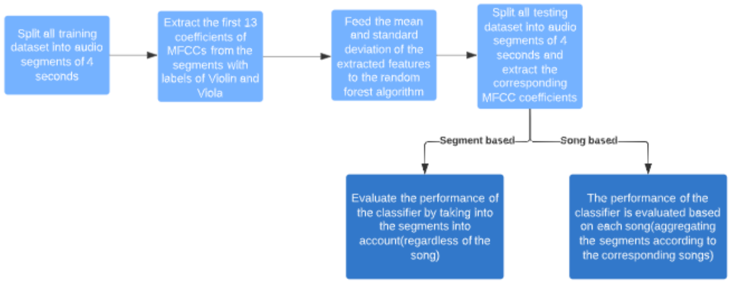


Figure 13: Flow diagram of using Frequency histogram

3.2.3 Mel-frequency cepstral coefficients(MFCC) features

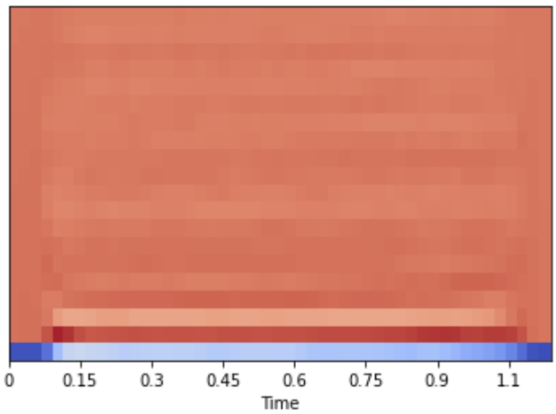


Figure 14: Visualization of MFCC feature of a violin playing A4 note

As shown by Figure 14 and 15, different musical instruments exude different MFCC features. Mel-frequency cepstral coefficients(MFCC) features are adopted as it is a dominant feature used in music information retrieval.

In order to get the MFCC features, a transformation from normal frequency scale to mel scale by applying the formula below needs to be done. It is due to the increasingly large intervals which are deemed to have equal pitch increments as listened by human ears.

The formula to derive mel scale with f denotes frequency is:

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right) = 1127 \ln\left(1 + \frac{f}{700}\right) \tag{1}$$

There are 40 filter channels in mel-frequency scale. The zero order coefficient output represents the mean power of the signal, the log-spaced

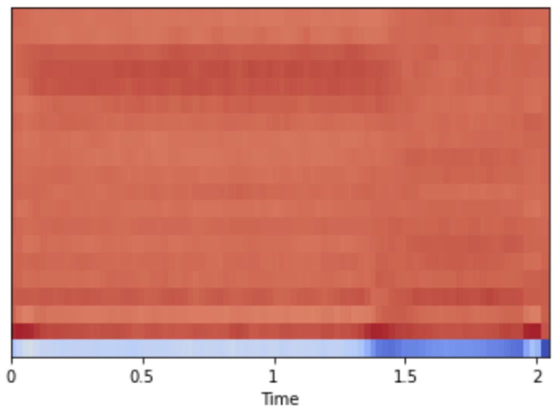


Figure 15: Visualization of MFCC feature of a viola playing A4 note

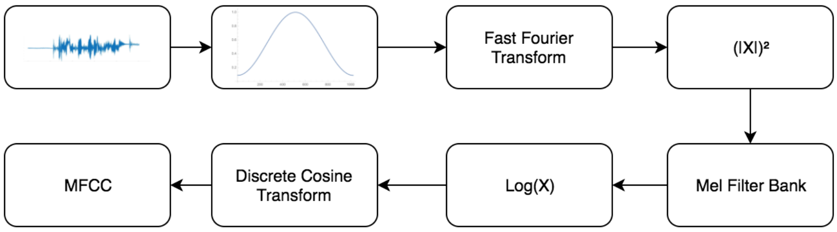


Figure 16: Flow of deriving MFCC features

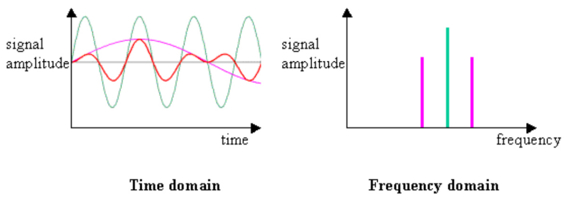


Figure 17: Transformation of audio signal from time domain to frequency domain via Fourier Transform

outputs represents the harmonics of the signal, while the twelve linearly spaced channels accounts for the spectral envelope. In a more specific context, the low order coefficient outputs shows the general spectral shape of source filter transfer function. The first order coefficient outputs contains the information about the distribution of spectral energy among high and low frequencies. The first thirteen coefficients are adopted. The corresponding means and standard deviations are used as the features to be fitted into the classifiers. In order to mitigate computational time and prevent any overfitting due to overly complex models, merely 13 cepstral coefficients are

Table 7: Evaluation metric for dataset(Solo and non-solo pieces) using Frequency Histogram

	Accuracy	Precision	Recall	F1-score
Baroque era	0.75	0.76	0.75	0.75
Classical era	0.83	0.83	0.83	0.83
Romantic era	0.83	0.88	0.83	0.83

Table 8: Evaluation metric for dataset(Solo pieces) using Frequency Histogram

	Accuracy	Precision	Recall	F1-score
Baroque era	1	1	1	1
Classical era	1	1	1	1
Romantic era	0.83	0.88	0.83	0.83

included as features.

3.3 Classifier and Evaluation

The classification algorithm used in this research is Random Forest, an ensemble classification method consists of a multitude of decision trees. The evaluation metrics included are accuracy, precision, recall, and F1-score.

4 Result

Table 9: Evaluation metric for dataset(Non-solo pieces) using Frequency Histogram

	Accuracy	Precision	Recall	F1-score
Baroque era	0.83	0.88	0.83	0.83
Classical era	0.67	0.67	0.67	0.67
Romantic era	0.83	0.88	0.83	0.83

Table 10: Evaluation metric for dataset(Solo pieces) using MFCCs

	Songs-based				Segments-based			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
Baroque era	1	1	1	1	0.98	0.98	0.98	0.98
Classical era	1	1	1	1	0.945	0.947	0.945	0.945
Romantic era	0.83	0.875	0.83	0.83	0.86	0.88	0.86	0.87

Although frequency histogram can be adopted to classify solo and non-solo pieces, it does not perform well on non-solo pieces(see in Table 7 and 8). However, the frequency histogram approach can achieve excellent results for Baroque and Modern solo pieces(see in Table 8). While this methodology cannot achieve same results for Romantic era, this is probably due to the reason that the overlapping notes in the repertoires of Violin and Viola of this era are more as compared to prior eras.

As shown by Table 10, MFCC can only be used to perform classification on solo pieces. However, this methodology can be capitalized to be tested on **song-based** and **segment-based** evaluation method. The evaluation performance on the basis of songs is better due to the underlying majority voting involved. From the results of MFCC implementation, it is clear that some 4-seconds segments of Violin and Viola performance exhibits very similar timbral features.

5 Conclusion

Although deep learning has been the state-of-the-art approach for classification tasks, but optimal features extracted can adopted to achieve decent classification with less computational cost without the need of Graphical Processing Unit(GPU). Since the classification of Violin and Viola is hard even for human experts, thereby creating a decent model to perform the corresponding classification task is definitely the main aim of this research. By taking the strings and the notes from repertoires of Violin and Viola, this work capitalizes on the **frequency histogram**, which is not adopted in prior research scholarships. The frequency histogram methodology can achieve great performance on solo pieces of Baroque and Classical eras. Moreover, MFCC is adopted as well as it is known to represent the timbral attributes of musical instruments well. Moreover,this work breaks the repertoires into 3 distinct eras, namely Baroque era, Classical era as well as Romantic era.

The difference of these two methodologies is frequency histogram can be tested on non-solo pieces while MFCC is extensible to other musical instruments by merely changing the dataset used. On the other hand, MFCC methodology is fine-tuned in this work by evaluating on the basis of songs and segments, which is different from existing research(evaluating on the basis of segments). The space requirements of frequency histogram approach is less as the variables need is 7 while the variables needed for MFCC approach is 26(means and standard deviations for the 13 coefficients)

To be discussed

References

[1] J.-J. Aucouturier and F. Pachet, "Scaling up music playlist generation," vol. 26, 02 2002, pp. 105 – 108 vol.1.

[2] T. Lidy and A. Rauber, "Evaluation of feature extractors and psycho-acoustic transformations for music genre classification." 01 2005, pp. 34–41.

[3] Y.-H. Tseng, "Content-based retrieval for music collections." 01 1999, pp. 176–182.

[4] P. Emanuele, L. Maurizio, and A. Giulio, "Musical instrument timbres classification with spectral features," *EURASIP Journal on Advances in Signal Processing*, vol. 2003, 01 2003.

[5] A. Diment, P. Rajan, T. Heittola, and T. Virtanen, "Modified group delay feature for musical instrument recognition," 2013.

[6] A. Eronen, "Comparison of features for musical instrument recognition," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, 2001, pp. 19–22.

[7] J. Marques and P. Moreno, "A study of musical instrument classification using gaussian mixture models and support vector machines," 09 1999.

[8] Z. Xiong, R. Radhakrishnan, A. Divakaran, and T. Huang, "Comparing mfcc and mpeg-7 audio features for feature extraction, maximum likelihood hmm and entropic prior hmm for sports audio classification," *Proceedings - ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 5, pp. 628–631, Sep. 2003, 2003 IEEE International Conference on Accoustics, Speech, and Signal Processing ; Conference date: 06-04-2003 Through 10-04-2003.

[9] I. Kaminskyj and T. Czaszejko, "Automatic recognition of isolated monophonic musical instrument sounds using knnc," *Journal of Intelligent Information Systems*, vol. 24, no. 2/3, pp. 199 – 221, 2005.

[10] P. Li, J. Qian, and T. Wang, "Automatic instrument recognition in polyphonic music using convolutional neural networks," *CoRR*, vol. abs/1511.05520, 2015. [Online]. Available: http://arxiv.org/abs/1511.05520

[11] E. Roche, "Of myths and Baroque music," *Early Music*, vol. 47, no. 1, pp. 132–135, 03 2019. [Online]. Available: https://doi.org/10.1093/em/caz011

[12] A. George, "Classical and Romantic chamber music for strings," *Early Music*, vol. XXIII, no. 2, pp. 341–348, 05 1995. [Online]. Available: https://doi.org/10.1093/earlyj/XXIII.2.341

[13] "Riffstation by riffstation - audio analysis tool application." [Online]. Available: https://www.kvraudio.com/product/riffstation-by-riffstation