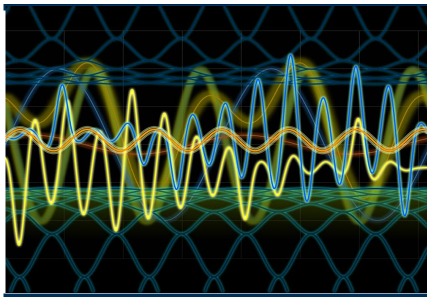
Proceedings of Meetings on Acoustics

Volume 49 http://acousticalsociety.org/



Fourth Vienna Talk on Music Acoustics
University of Music and Performing Arts

Vienna, Austria
11-14 September 2022

Musical Acoustics: Poster Session, Set. 12

Identification of violin timbre by neural network using acoustic features

Masao Yokoyama and Yuya Ishigaki
Department of Information Science, Meisei Daigaku, Hino, Tokyo, 191-8506, JAPAN; masao.yokoyama@meisei-u.ac.jp; 22DJ002@stu.meisei-u.ac.jp

The timbre of violins is identified using machine learning, and a computer program is developed for the neural network using Python and Keras libraries. The 21 violins recorded include old Italian violins made by Stradivari and contemporary violins. The training and test data use the spectrum envelope and Mel-frequency cepstrum coefficients (MFCC). The accuracy of the identification test in the case of open strings is greater than 90 %. Furthermore, experiments that predict similarity in timbre of an unknown violin to that of trained violins are presented.

1. INTRODUCTION

Some professional violinists and dealers can distinguish differences in the timbre of violins, such as differences in the country of the product and differences in violin makers. However, the ability to distinguish the timbre of a violin is sensitive and intuitive; it is limited to professionals who have the opportunity to observe many antique violins, such as Stradivari and Guarneri del Gesu. Thus, non-professionals have difficulty in distinguishing differences in the characteristics of antique violins.

Several studies have analyzed the timbre of violins. Using spectrum analysis, Buen analyzed the difference in the acoustic features of 30 violins, including Stradivari and Guarneri del Gesu.[1,2] Old and modern violins have been compared by a collaboration between the university and the violin museum in Cremona.[3] Similarly, in Taiwan, Tai and Chung[4] studied the formant analyses of Stradivari and Amati using the cepstrum (LPC method, linear prediction coding method), with collaboration with the Chi Mei Museum. Nagyvary compared the sound of Guarneri del Gesu and vowel sounds by formant analysis.[5] For a musical instrument with a resonance body, the acoustic features can be analyzed by observing the spectral envelope and formant.[6]

This study attempts to determine whether a computer can identify differences in the timbre of violins like violinists and dealers can. In recent years, machine learning techniques, such as deep learning, have been rapidly developed. By learning considerable sound data, the identification of a timbre is possible using machine learning techniques. Moreover, an antique violin can be appraised and a more realistic sound can be synthesized using generative models of deep learning. For example, the sound quality of electric instruments, such as an electric violin, digital piano, and MIDI sound, is expected to improve and be more realistic. In this study, we recorded 21 violins, from old Italian violins made by Stradivari to contemporary violins. Additionally, we calculated the spectrum envelope and Mel-frequency cepstrum coefficients (MFCC) for the training and test data. Subsequently, we evaluated whether artificial intelligence can distinguish the timbre of different violins (e.g., Stradivari's versus other violins) and identify the violin maker.

2. RECORDING AND MACHINE LEARNING

RECORDING OF VIOLIN SOUND

The violin sounds listed in Table 1 were recorded. The violins used for recording were selected from the old Italian violins of the 17th century to the Japanese bland-new commercial violins. We recorded the sound that a violinist played on the open strings without any expression (E5, A4, D4, G3, long-tone of approximately 4 s) and the musical piece "Meditation from Thaïs" with musical expression (including vibrato, dynamics, diminuendo, and crescendo). This musical piece was in a slow tempo with several long notes. Moreover, the pitch varied over a wide range, from A3 (220 Hz) to F#6 (1480 Hz), which is advantageous for learning various sound data. The violinist played the open strings and musical pieces twice for each violin.

As for the recording conditions, FFT analyzer (Oros NV Gate OR30 series) and the ICP 1/4-inch microphone was used for recording. The microphone was set at approximately 10 cm above the violin bridge. The microphone's frequency response was 20–20 kHz, the dynamic range was 30–143 dB, and the sampling frequency was 51.2 k/s. The sound wave was saved as a wav 16-bit PCM format. The sounds of the violins were recorded in a small rehearsal room in a violin shop. The reverberations and echoes in the room were very small.

Table 1. Violins for recording and training data

Violin maker (country)	**Year**
Catenali (Italy)	1690 ca
A.Stradivari (Italy)	1698
Pietro Guarneri (Italy)	1700 ca
Santo Serafin (Italy)	1700 ca
Gragnani (Italy)	1760
Balestrieri (Italy)	1780

Pressenda (Italy)	1838
Fabris (Italy)	1870
Scarampella (Italy)	1907
Fagnola (Italy)	1923
Genovese (Italy)	1927
Michetti (Italy)	1929
Guerra (Italy)	1941
Bisiacchi (Italy)	1953
Garinberti (Italy)	1967
Contemporary violin middle-class A (Japan)	2015
Contemporary violin middle-class B (Japan)	2015
Contemporary violin Economic A (Japan)	2015
Contemporary violin Economic A (Japan)	2015
Contemporary violin Stradivari Copy (Japan)	2015
Contemporary violin Guarneri Del Gesu Copy (Japan)	2015

PROGRAM FOR MACHINE LEARNING

Figure 1 illustrates the methodology used in this study. The acoustic features, spectrum envelope, and MFCC of the recorded sound data were calculated using a computer program. Supervised neural network training was performed using a pair of acoustic features and violin makers; herein, the input was the acoustic features, and the output was the violin maker. Finally, the trained network was run on the test data to validate the accuracy and calculate the similarity for each violin.

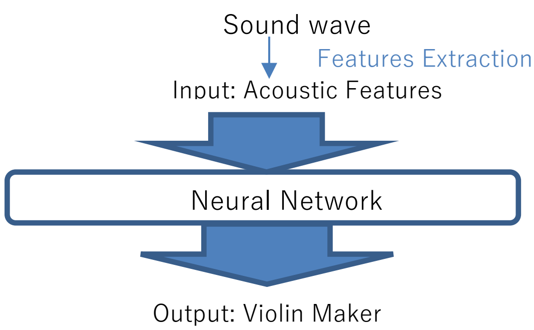


Figure 1. Schematic overview of the procedure

The program to calculate the acoustic features and execute machine learning was written using Python and the Keras library, which is the front end of TensorFlow. The program consists of numerous functions: division of sound waves, calculation of the spectrum envelope and MFCC, and execution of machine learning and evaluation. The sound wave where the sound of a violin is recorded was cut into 0.06 s segments; subsequently, this short sound wave data was multiplied with a window function, the Hanning window.

The program calculates the spectrum envelope using the cepstrum analysis method[7] and saves the spectrum envelope with the label that identifies the violin in a CSV file. Herein, the power spectrum was calculated using a digital Fourier transfer (DFT). Furthermore, the logarithmic spectrum provided by the logarithmic conversion was performed with inverse DFT, and the axis returned to the time domain, after which the cepstrum was obtained. After the low-level part of the cepstrum was filtered (liftering) and DFT was performed, the axis was shifted to the frequency domain again, and the spectrum envelope was obtained.

If the cepstral coefficient is larger, the shape of the spectrum envelope becomes more complex. In this study, the sampling number of the DFT was set to 1024 points, and the cepstrum coefficients were set to 20–80. The spectrum envelope data were divided into two subsets for the training and testing of the neural network. The training dataset was selected randomly from the entire dataset of all violins. The percentage of testing data set against the entire dataset was 5 %, and the remaining 95 % of the entire dataset was used for the training data set. The label for identification was assigned to each of the 21 violins and four strings; thus, 84 labels were assigned to distinguish the timbre. The neural network was a fully connected four-layer network. The number of inputs was 1024, and the outputs of the 2nd and 3rd layers were 512. The number of outputs in the 4th layer was the same as the number of labels. The activation function was a ReLU function. The learning rate was 0.1, and the dropout ratio was 0.2.

3. VALIDATION

A. IDENTIFICATION OF TIMBRE

The result of the experiment identifying the violin using the neural network trained with only open-string data is depicted as the blue line in Fig. 2. This result depicts the accuracy with which the program correctly identifies the violin and its string, where C is the cepstral coefficient. The number of training data was approximately 14000, and the number of evaluation data was approximately 700. Evidently, accuracy of more than 95 % was obtained in the case where C was 60 and 80. The reason for the good accuracy was attributed to the fact that the sound wave of the open string was approximately periodic.

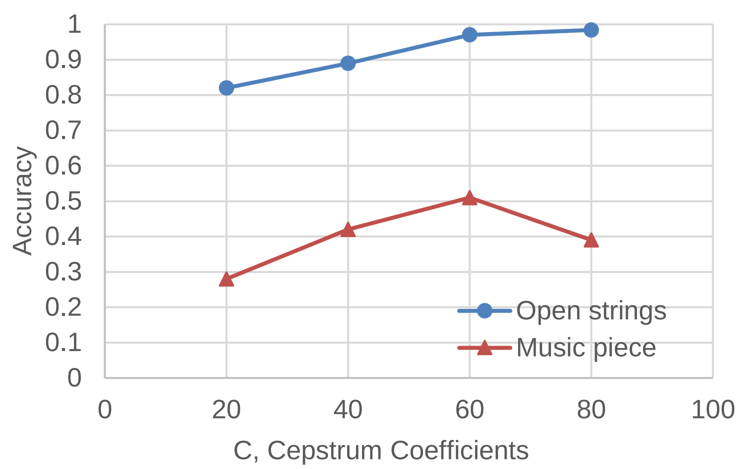
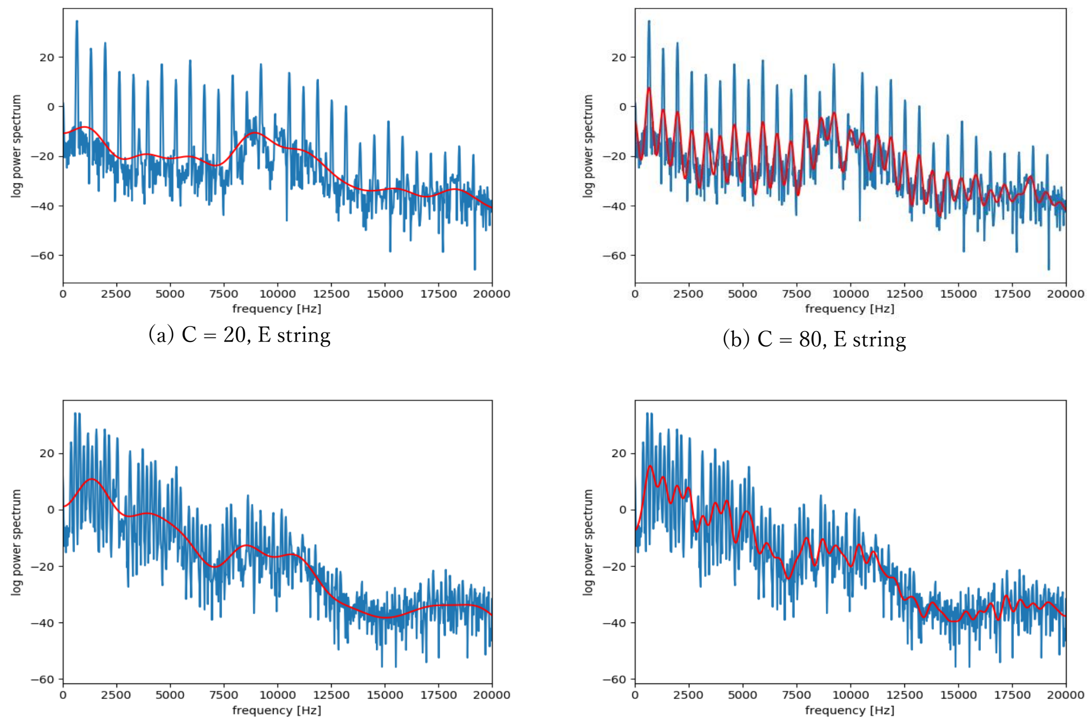


Figure 2. Accuracy of identification of violin (average). In the case of open strings, the accuracy is more than 95 %. But, in the case of a music piece, it is approximately 50 %

The spectrum envelopes of strings E and G at different C values are shown in Fig. 3. Evidently, the spectrum envelope traced the peaks of the spectrum when C was low (Figs. 3 (a) and (c)). However, the spectrum envelope became wavy when C was high, as shown in Fig.3 (b) and (d). In particular, in Fig. 3 (b), the peaks of the spectrum envelope were almost identical to those of the spectrum. Thus, when the pitch of the sound of the evaluation data is high, an appropriate envelope curve is not obtained, and the accuracy is reduced. To solve this problem, further investigation and tuning of C are required.



(a) C = 20, E string

(b) C = 80, E string

(c) C = 20, G string

(d) C = 80, G string

Figure 3. Spectrum envelopes with different Cepstrum coefficients, C (open E and G string, Violin: Stradivari 1698).

Next, identification using the dataset of the music piece was performed. Approximately 40 s of sound were used for the datasets from the beginning of the music theme. The number of training data was approximately 70000, and the number of evaluation data was approximately 3500. The accuracy is depicted as an orange line in Fig. 2. Because of the changes in pitch, dynamics, vibrato, and other musical expressions in the performance of the musical piece, the spectrum envelope changed much more variously than that of the open strings. C was set at 60, where the envelope was moderately generated. The accuracy of violin identification was approximately 50–60 %. The accuracy of the music piece was lower than that of the open strings because of the sound wave's non-constancy.

Figure 4 shows a comparison of recognition accuracy depending on the nature of the sound. The accuracy was high when the neural network was trained and tested using open-string data only. The accuracy was approximately 90 %. This is because the sound wave of the open string was approximately periodic. However, when the training data was the music piece with expression, the accuracy became low (middle of the graph in Fig. 4, without string distinction) because the music piece contains a complex set of parameters, including vibrato and dynamic changes. In addition, in the case of the music piece, when the violins and strings (E, A, D, and G strings) were identified by calculating F0, the pitch range was recognized, and different labels were assigned to the training data for each combination of violin and string, the accuracy increased slightly (right side of the graph in Fig. 4). Thus, the accuracy depends on the pitch and selection of cepstrum ecoefficiency.

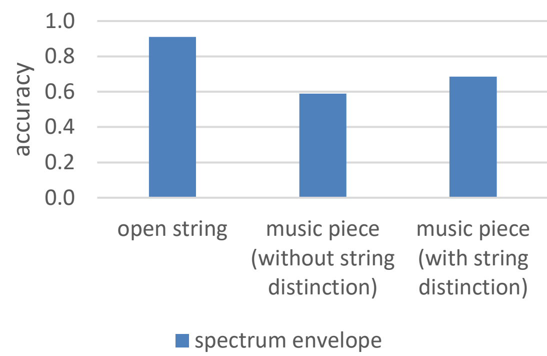


Figure 4. Accuracy based on the nature of sound

Figure 5 compares accuracies in acoustic features (spectrum envelope, Mel-spectrum, and MFCC). This result shows that MFCC is the best for identifying violin sounds.

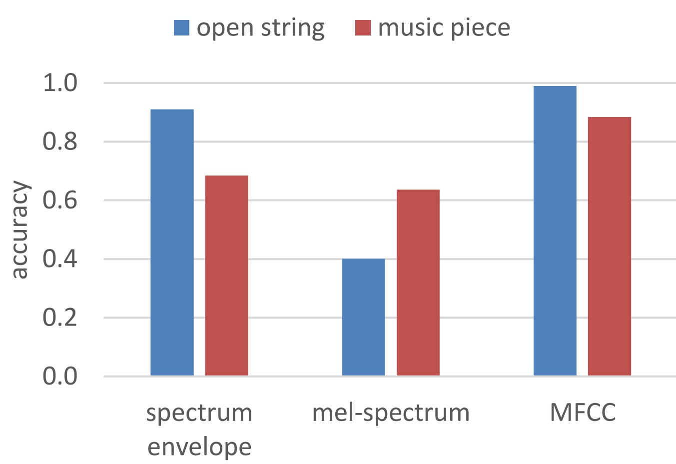


Figure 5. Comparison in difference with acoustic features

B. APPLICATION TO QUANTIFICATION OF TIMBRE SIMILARITY

Using the identification system of the neural network, an application for quantitatively predicting the similarity of timbre between two violins is discussed in this section.

When purchasing a violin, the question often asked is, "Which violin sounds similar to the Stradivarius the most?" In this situation, having a tool that can calculate and visualize the similarity in timbre among violins quantitatively using AI may be helpful and convenient. Therefore, we tested whether our system can quantify the timbre and predict the similarity between an unknown violin (you want to buy) and a famous violin.

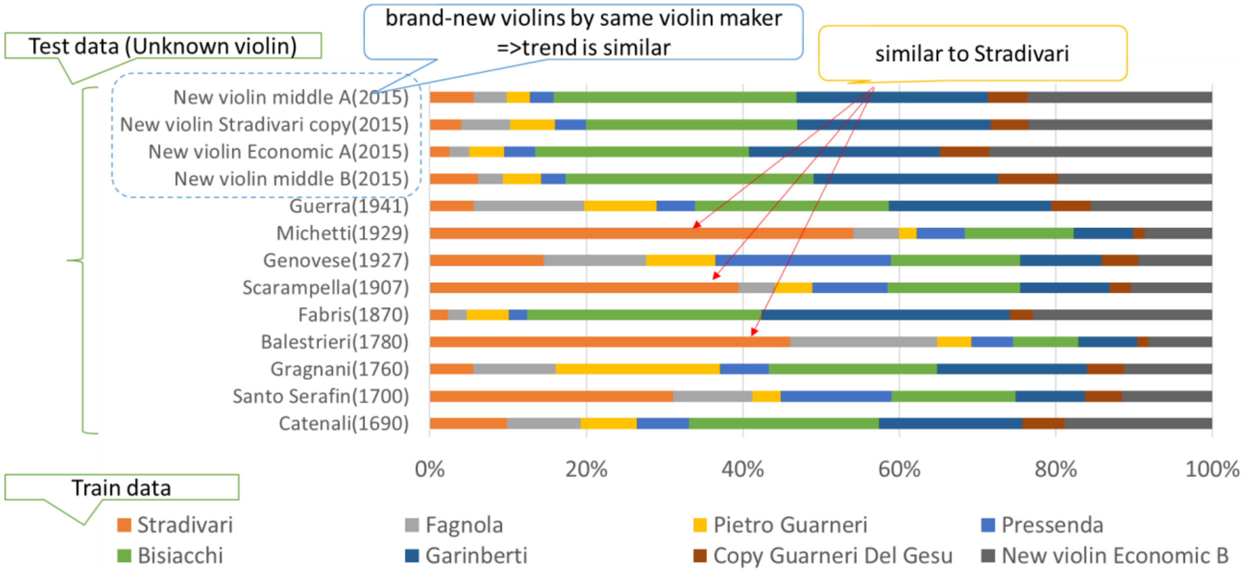


Figure 6. Percentages of the timbre similarity of test violins against eight violins (famous violins, such as Stradivari and Fagnola), Dataset: Performance of a music piece, MFCC. Sant Seraphin is similar to Stradivari at approximately 42 %, and Michetti is similar to Stradivari at approximately 54 %. In contrast, the timbre of bland-new violins resembles each other, similar to Garinberti, Bisiacchi, and Bland-new economic B

The probabilities which our neural network predicted are shown in Fig. 6. The eight violins (legend under the graph) selected from Table 1, including Stradivari and Fagnola, and their MFCC, were used to train the neural network. The other violins in the vertical axis are violins, assumed as unknown violins, to predict the similarity.

For example, our program predicted that Michetti is similar to the Stradivari by 54 %, and Sant Seraphin is similar to Stradivari by 42 %. In contrast, the timbre of the bland-new violins made in Japanese factories resemble each other and are similar to the timbre of modern violins, such as Bisiacchi, Garinberti, and bland-new violin B (made by the same violin factory).

However, we used one violin for each violin maker in this experiment; hence, we cannot conclude definitively that the trained model of the network can express the general characteristics of a violin maker's timbre. More recording data per violin maker is required for correct appraisal.

4. CONCLUSION

Experiments to identify the timbre of violins using a neural network were performed using the sound data from 21 violins. The possibility of quantifying the similarity of violin timbre and identifying the violin maker using a neural network trained by acoustic features, such as the spectrum envelope and MFCC, was demonstrated. In future, we may be able to develop an appraisal machine to authenticate Stradivari's violin and synthesize more realistic sounds for electric instruments, such as the sound of a digital piano and MIDI sounds.

ACKNOWLEDGMENTS

The authors would like to thank Bunkyo-Gakki Co. Ltd. for providing violins. This study was supported by JSPS KAKENHI (grant number 21K11969).

REFERENCES

[1] A. Buen, “Comparing the sound of golden age and modern violins: Long-time-average spectra,” *VSA papers*, Vol. 1 (1), pp. 51-74 (2005).

[2] A. Buen, “What is old Italian timbre?,” *Proceedings of the Second Vienna Talk (VITA), University of Music and Performing Arts, Vienna: VITA*, pp. 19-21 (2010).

[3] F. Setragno, M. Zanoni, F. Antonacci and A. Sarti,. “Feature-Based Timbral Characterization of Historical and Modern Violins,” *International Symposium on Musical Acoustics*, pp. 90-93 (2017).

[4] H. C. Tai and D. T. Chung, “Stradivari violins exhibit formant frequencies resembling vowels produced by females,” *Savart Journal*, Vol. 1 (2) (2012).

[5] J. Nagyvary, “A comparative study of power spectra and vowels in Guarneri Violins and Operatic Singing,” *Savart Journal*, Vol. 1 (3) (2013).

[6] M. Yokoyama, “Possibility of distinction of violin timbre by spectral envelope,” Applied Acoustics, 157, 107006, (2020).

[7] S. Furui, "Cepstral analysis technique for automatic speaker verification," IEEE Transactions on Acoustics, Speech, and Signal Processing Vol. 29. 2, pp. 254-272 (1981).