

ニューラルネットワークを用いたヴァイオリンの音色の類似性識別

横山 真男[○](明星大学)

Similarity identification of timbre of the violin using the neural network

Masao YOKOYAMA

ABSTRACT

Identification of the timbre of violins was performed using a machine learning technique. The machine learning computer program was developed using Python and Keras library. The number of violins tested is 21, which contains old Italian violins such as Stradivari. As for acoustic features, the spectrum envelope calculated by the cepstrum method was used for the training data and the test data. As the result of identification experiment, in the case of open strings an accuracy of more than 93% was obtained. Furthermore, an experiment that analyzes the similarity of timbre between an unknown violin and trained violins was shown.

Keywords: Neural network, Timbre, Violin, Cepstrum, Spectrum envelope

1. はじめに

本研究は、人間のもつ感覚的な楽器音の音色の違いの識別が、計算機でも模擬できるのかどうかチャレンジするものである。近年盛んに研究開発されている強化学習を用いて、多量の音響データを学習してヴァイオリンの音色の違いを区別させようというものである。ヴァイオリンという弦楽器においては、昔からストラディバリのような高額な楽器の音色の良し悪しについて議論が行われてきた¹⁻³⁾。また、楽器をよく知る演奏者や鑑定家は、有名な楽器製作者が作ったヴァイオリンに対して感覚的にその楽器製作者らしい特徴的な音色の区別ができるという。こういった感覚的な話はもちろん科学的な根拠がほぼないに等しく、またプロの中でも多くの楽器を手にしたことのある限られた専門職の人たちだからこそ分かる繊細で直観的なレベルの話であり、ノンプロにとっては区別するのは難しい。これが可能になると、いわゆる楽器の鑑定が計算機でできるようになるだけでなく、生成モデルを使ってよりリアルなサウンドの合成ができるようになり、エレキヴァイオリンやデジタルピアノのような電子楽器の音色を、よりリアルな音質への改善が期待できる。

2. 音源の収集とケプストラム分析のプログラム

2.1 演奏音の録音

ヴァイオリンの音の収集については、各ヴァイオリンのE線、A線、D線、G線の開放弦と、楽曲の演奏音として「タイスの瞑想曲」を録音した。この楽曲はゆったりとした曲想で、二分音符や全音符が多く、また音高が上下し音域も広い(A3,220Hz-F#6,1480Hz)ため学習には都合が良い。

合が良い。

Table 1に本分析に使用したヴァイオリンを示す。17世紀のオールドイタリアンから20世紀のモダン楽器、また日本の新作の量産品まで幅広く収録した。

Table 1. Violins for training data and test 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
New violin middle class A (Japan)	2015
New violin middle class B (Japan)	2015
New violin Economic A (Japan)	2015
New violin Economic A (Japan)	2015
New violin Stradivari Copy (Japan)	2015
New violin Guarneri Del Gesu Copy (Japan)	2015

演奏者はプロの奏者 1 名である。楽曲の演奏については f や p などの強弱やビブラートなどの表現もつけて演奏してもらっている。録音と分析には FFT アナライザを使用し、マイクロホンは 1/4 インチ、周波数応答は 20Hz~20kHz(± 2dB) で、ダイナミックレンジは 30dBA~143dB である。マイクは演奏者の持つヴァイオリンの f 字孔から約 5cm 離れた位置に固定し録音を行った。

2.2 ケプストラム分析によるスペクトル包絡と学習プログラム

スペクトル包絡は音声信号処理における音響パラメータとして重要な情報である。音声信号からパワースペクトルを DFT で求めると、音素に応じてある周波数帯に強いパワーが現れ、その包絡線を求めると局所的なピークがいくつか観測される。音声信号処理の分野ではフォルマント解析などの様に母音の認識に利用されるが、近年では楽器の音色への適用も試みられている^{4,5)}。楽器の本研究では、ケプストラム分析⁶⁾によるスペクトル包絡を学習データとして用いた。

ケプストラム分析および深層学習をするプログラムは、Python および TensorFlow をバックエンドとするライブラリの Keras を用いた。それぞれ、波形分割、ケプストラム分析、データセット生成、学習、評価の各処理を行うものである。以下に、プログラムの処理概要を以下に示す。録音音源を 0.03 sec ずつ切り取り、ハニング窓をかけて多くの短区間の音データを得る。得られた音データを、ケプストラム分析にかけ^[7]、スペクトル包絡を抽出し、ヴァイオリンを識別するラベルの値と共に csv ファイルに格納する。

そして離散フーリエ変換を行うと、時間領域であった波形から周波数領域に移行し、スペクトルが得られる。更に対数変換によって得られた対数スペクトルを逆離散フーリエ変換すると、再び時間領域に戻り、ケプストラムが得られる。ケプストラムの低次部分をフィルタにかけて(リフタリング)離散フーリエ変換すると、再度周波数領域に移行してスペクトル包絡が得られる。設定したケプストラム次数の値が大きいくほど得られるスペクトル包絡の形が複雑になる。本研究では FFT のサンプル数を 2048 点、ケプストラム次数を 20-80 に設定した。

ケプストラム分析したデータは学習用データセットと評価用データセットに分ける。この時、評価用データセットは学習用データセットとは同じ楽器であるが違う区間を間欠的に取り出して使っている。すなわち、一つのヴァイオリンから得た全ケプストラムデータのうち 5% のデータを間欠的に選び評価用データとし、残りを学習用データにした。

ラベルは、開放弦についてはヴァイオリン 21×弦 4 の 84 のラベルを振っている。また、楽曲については、音高一つずつに振るのは困難であるので今回はヴァイオリン毎で 21 のラベルを振っている。

ネットワークは全結合ニューラルネットワーク 3 層で、入力数は 1024、1 層と 2 層の出力数は 512、3 層の出力数はラベル数と同じで、活性化関数は ReLU 関数を用いている。学習率は 0.1 でドロップアウトは 0.2 に設定した。

3. 実験結果

3.1 開放弦のみのデータによる識別率

開放弦の音のみで楽器および弦を識別した結果を Fig.1 に示す。評価データがどのヴァイオリンの何弦であるかを当てるものであり、その正答数を識別率 (Accuracy) としケプストラム次数 C (Cepstrum Coefficient) で比較した結果である。学習用データ数は約 14000、評価用データ数はおよそ 700 で、各 C につき 3 回の計算結果を平均したものである。ケプストラム次数が 60 あたりで約 93% の精度が得られた。開放弦では後に示すように楽曲に比べ、音の波形がほぼ定常であるため比較的良好な精度が出ているものと思われる。

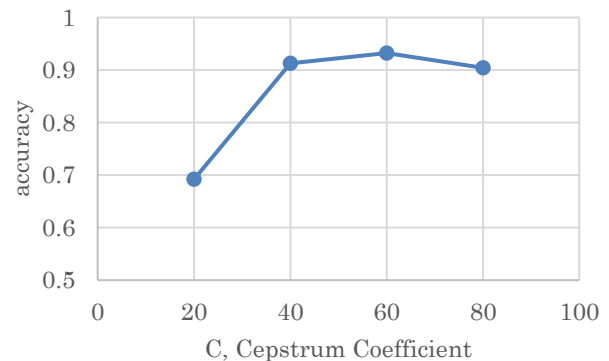


Fig.1 Difference of accuracy in the cepstrum coefficient.

さらに詳細な分析のために弦毎に比較した結果を Fig.2 に示す。E 線では $C=80$ が最もよい識別率であるが他の弦では $C=60$ が良く、また G 線では $C=80$ の値が低い。結果として、60 が最も良く次いで 40 が設定値として良いことがこのグラフから分かる。

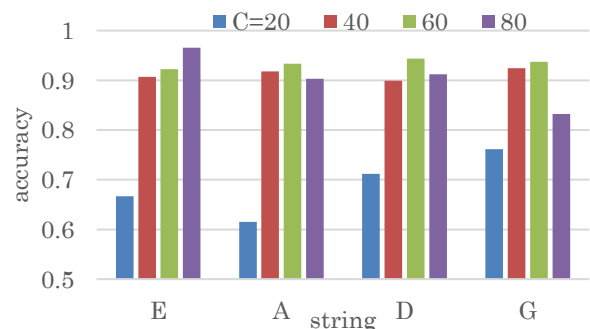
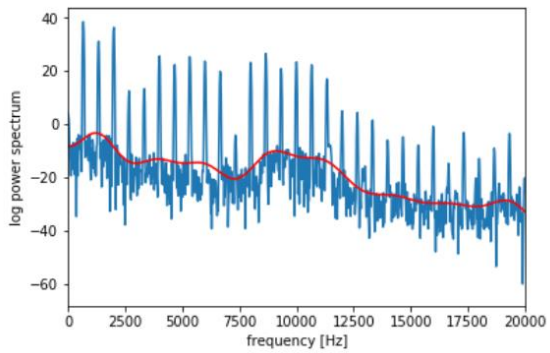
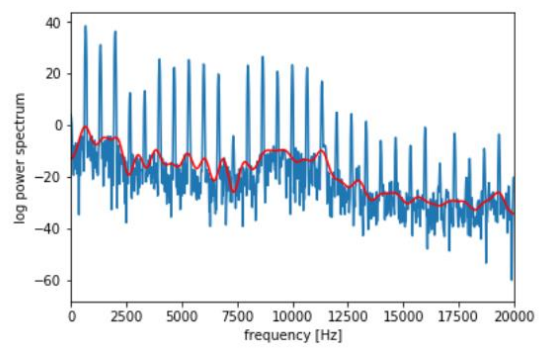


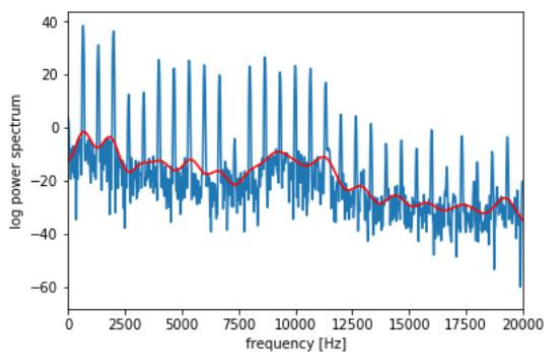
Fig.2 Difference of accuracy in the strings of violin.



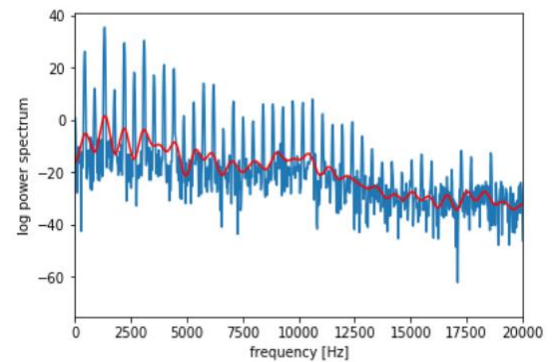
(a) $C = 20$, E string



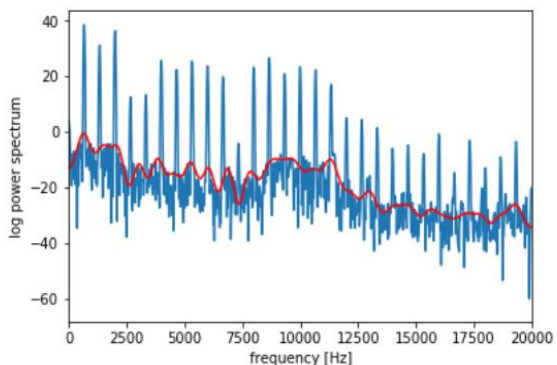
(a) $C = 60$, E string



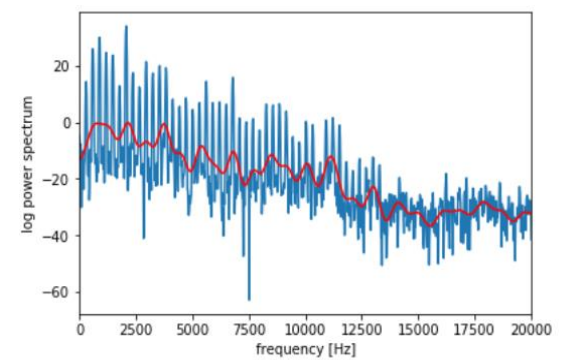
(b) $C = 40$, E string



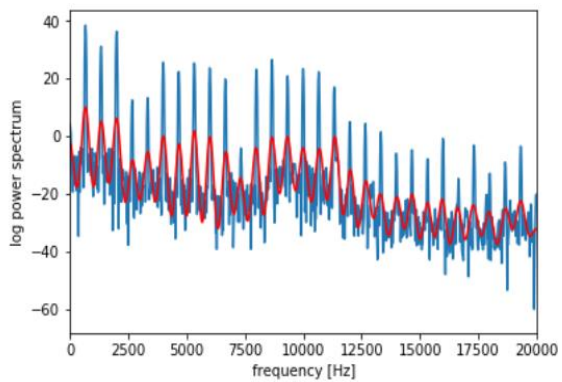
(b) $C = 60$, A string



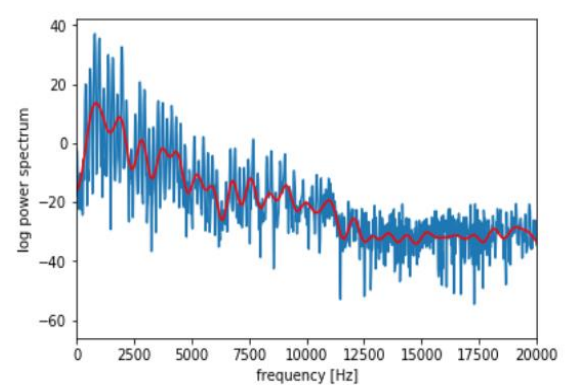
(c) $C = 60$, E string



(c) $C = 60$, D string



(d) $C = 80$, E string



(d) $C = 60$, G string

Fig.3 Difference of spectral envelope in Cepstrum Coefficients, C . (Open E string, Stradivari 1698)

Fig.4 Difference of spectral envelope in string (Open E, A, D, G string, Stradivari 1698)

また、Fig.3 および Fig.4 にケプストラム次数 C の違いによるスペクトル包絡の表現の違いを示す。一番高い音の開放弦である E 線のスペクトル包絡は (Fig.3) $C=80$ では細かすぎて概形というよりスペクトルのピークそのものをなぞるようになり、逆に、 $C=20$ では大まかすぎて包絡線がスペクトルの概形をとらえきれていない。一方、Fig.4 のように $C=60$ ではスペクトルの概形を程よくなぞっているのが見て取れ、これが識別率の値に影響しているものと思われる。

3.2 演奏音のデータによる認識率

楽曲の音源による学習用データセットで認識実験を行った。開放弦に比べ、音高の変化、ビブラート、音量等によりスペクトラムは大きく変動する。楽曲のうち最初のテーマの部分で 40 秒ほどの音源をデータとして使用した。音域は、下は G 線の A3 から上は E 線の F#6 である。各ヴァイオリンを合わせたデータ全体の数は、学習用データセットが約 70000、評価用データセットが約 3500 である。

結果は、楽器の学習順序などを変えて行ったが、識別率はおよそ 54% から 59% の間であった。ケプストラム次数については開放弦で程よく包絡線が得られる $C=60$ が同様に最も識別率が良かった。やはり、開放弦に比べ非定常性が大きいいため識別の精度は落ちている。スペクトル包絡等の音響モデルの検討 (LPC, MFCC など) や、ニューラルネットワークのモデルなどのチューニングが今後の課題である。

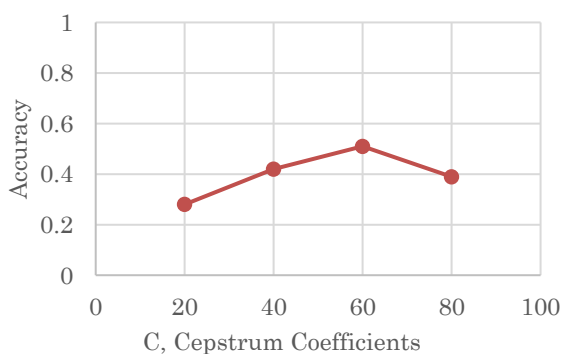


Fig.5 Difference of accuracy in the cepstrum coefficient by music piece.

3.3 未知のヴァイオリンの音色傾向分析

上記のヴァイオリンの学習プログラムを使ったアプリケーション例を示す。例として、未知のヴァイオリン (学習用データセットにないヴァイオリン) が、どのヴァイオリンに近いかを見積もる実験を示す。例えば、ヴァイオリンを買おうとしている購入者が、そのヴァイオリンがどんな有名な楽器に似ているのか、ストラディバリに近いのかグァルネリに近いのかが数値的にわかり、楽器の選択のひとつの指標になる。

Fig.3 に実験結果の例を示すが、表 1 の中より 7 挺のヴァイオリンを学習用データセットとし (横軸)、新作の 6 挺のヴァイオリン (縦軸) を評価データセットに使った場合である。評価した未知の 6 挺が、学習した 7 つのヴァイオリンにどれだけ似ているかを割合で表している。同じメーカーの工場製新作ヴァイオリンが似たような音色であったことを示し、20 世紀のモダンヴァイオリンである Bisiacchi や Garinberti、また P.Guarneri により近いことを示している。

しかし、今回のデータはあくまでも各製作者の楽器が 1 挺ずつの学習データであるため、製作者のとある楽器に似ているか否かという結果であり、その製作者が作る全ヴァイオリンの傾向を示しているわけではない。各製作者に対し多くの楽器を学習させないと、その製作者の作るヴァイオリンの音色を示す判定にはならない。

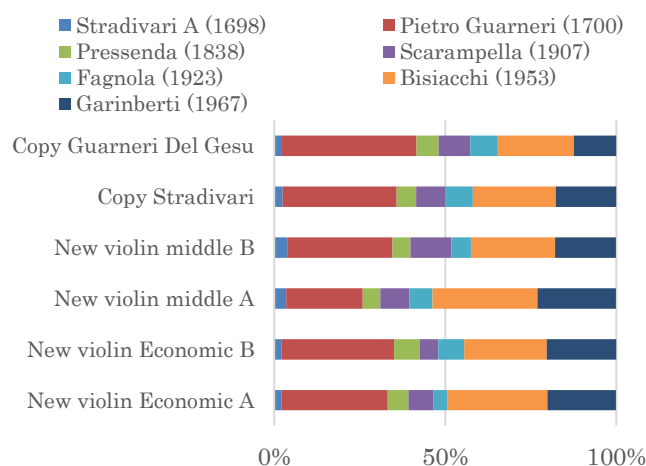


Fig.6 An application of our deep learning program. Percentages of timbre similarity against 8 violins (Stradivari, Fagnola, ...). For example, this chart shows new violins are similar to old violin, P.Guarneri, and modern violins such as Bisiacchi or Garinberti.

4. おわりに

イタリアの名器から国産の量産楽器まで 21 挺のヴァイオリンの音源を用いて、深層学習による音色識別の方法について述べた。準定常的な開放弦の音源では 95% の識別ができたが、楽曲の演奏では 54%~59% の識別にとどまった。また、本研究で用いたプログラムのアプリケーションの例として、学習データセットにないヴァイオリンがどの程度、学習データセットにあるヴァイオリンに音色が似ているかの割合を算出できるかについて示した。今後の課題として、学習用のヴァイオリン音源の収集、音響モデルの導出方法の違いの検討、ニューラルネットワークのモデルの検討などがある。

参考文献

- 1) Buen, A. (2005). Comparing the sound of golden age and modern violins: Long-time-average spectra. *VSA papers*, 1(1), pp. 51-74.
- 2) Buen, A. (2010). WHAT IS OLD ITALIAN TIMBRE?. *Proceedings of the Second Vienna Talk (VITA)*, University of Music and Performing Arts, Vienna: VITA.
- 3) Setragno, F., Zaroni, M., Antonacci, F., and Sarti, A. (2017). Feature-Based Timbral Characterization of Historical and Modern Violins. In *International Symposium on Musical Acoustics*, pp. 90-93.
- 4) Tai, H. C., and Chung, D. T. (2012). Stradivari violins exhibit formant frequencies resembling vowels produced by females. *Savart Journal*, 1(2).
- 5) Nagyvary, J. (2013). A comparative study of power spectra and vowels in Guarneri Violins and Operatic Singing.
- 6) Furui, S. (1981) "Cepstral analysis technique for automatic speaker verification." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 29.2, pp. 254-272.