

深層学習によるヴァイオリン製作者識別の試み

横山真男^{†1}, 八代月光^{†1}, 植木一也^{†1}

概要: 深層学習を用いて、演奏音からそのヴァイオリン製作者を当てる楽器識別および未知の楽器がどの製作者のヴァイオリンに近い音色分析を行うプログラムを開発した。音響特徴量としてケプストラム法によるスペクトル包絡を深層学習の学習およびテストデータに用いた。ヴァイオリンは、ストラディバリをはじめとするオールドイタリアンの名器からモダン、国産の新作楽器まで 21 本で評価実験を行った。開放弦の演奏音を学習させたところおよそ 90%以上の割合で識別ができ、楽曲の演奏音では約 60%弱の正答率が得られた。また、未知の楽器が学習したどの楽器に類似しているかといった傾向分析についても試みた。

キーワード: ヴァイオリン, 音色, ケプストラム, 深層学習, 人工知能

An attempt of recognition of violin makers by deep learning

MASAO YOKOYAMA^{†1} TSUKIHIRO YASHIRO^{†1}
KAZUYA UEKI^{†1}

Abstract: A recognition of violin makers and an analysis of violin timbre were performed by deep learning program. The recognition program detects which violin maker the recorded sound data belongs to and which trained violin an unknown violin is similar to. As the acoustic features, the spectral envelopes calculated by the cepstrum method are used for the training data and the test data. The number of violins played in this study is 21, which contains old Italian violins (including Stradivari's), modern violins and new ones made in Japan. From our experiments, in case of open strings, the accuracy of recognition was more than 90 %, and in case of performing a piece of music, that was approximately 60 %. Furthermore, the trend analysis chart which depicts the similarity between an unknown violin's timbre and the trained violins is shown in this paper.

Keywords: Violin, Timbre, Cepstrum method, Deep Learning, Artificial Intelligence

1. はじめに

楽器の音色は一本一本違うが、非常によく楽器を知る演奏者や鑑定家においては、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を言い当てることができる。つまり、イタリアの名匠の違いであったり、ドイツ製ヴァイオリンとフランス製ヴァイオリン等との違いが分かったりする。例えば、Pressenda (1777-1854, Trino)とその弟子の Fagnola (1865-1939, Torino)などは、音色のキャラクターの違いが語られることがしばしばある。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直観的なレベルの話である。

ヴァイオリン製作者の違いによる音響特徴量の研究については、Buen が Stradivari と Guarneri del Gesu を実に 30 挺という多くのヴァイオリン（これだけ名器を集めるのはかなり至難である）について、音色の違いをスペクトル解析で分析している[1-2]。また、ミラノ工科大学(Milano Politecnico)とクレモナのヴァイオリン博物館 (Museo del Violino) の共同による古今の楽器比較もある[3]。同様に、

博物館との共同では台湾の Chi Mei Musium と Tai & Chung[4]が LPC ケプストラムで Stradivari や Amati などのフォルマント分析をしている（余談であるが日本でもこのような楽器を多く所有している財団と大学との連携があると望ましい）。フォルマントによる比較でいえば、Guarneri と母音の比較を Nagyvary が行った[5]。これらのように、楽器も共鳴体を持っているので人間の声道特性ではないがフォルマントで分析できる可能性もあり著者らも取り組んでいる[6]。

本研究は、そのような人間のもつ感覚的な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深層学習 (Deep Learning) を用いて、多量の音響データを学習してヴァイオリン製作者の違いを区別させようというものである。これが可能になると、俗にいう楽器の鑑定が計算機でできるようになるだけでなく、また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、例えばストラディバリのような音色に合成することが可能になり、よりリアルな音質への改善が期待できる。

^{†1} 明星大学
Meisei University

2. 実験手順

2.1 ヴァイオリン演奏音の録音

演奏音の収録については、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、二分音符や全音符などが多く、また音高が上下し音域も広い (A3,220Hz - F#6,1480Hz)、二長調なので楽器を鳴らしやすいというメリットから選択した。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。

録音機器は FFT アナライザを使用し、マイクロホンは $\frac{1}{4}$ インチ、周波数応答は 20Hz~20kHz(± 2 dB)で、ダイナミックレンジは 30dB~143dB である。録音を行った部屋は、広さ 4 畳ほどの残響の少ない部屋を使用した。部屋の中心あたりに演奏者の方に立ってもらい、演奏者の正面に楽譜とマイクスタンドを設置した。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。奏者には極力動かないで弾くように指示した。

本分析に使用したヴァイオリンは 21 挺で Table 1 のとおり 17 世紀末から数年前まで幅広く収録した。新作は日本製の量産品 (Economy / Middle class) とマスター製作のものである。

Table 1. Violins for training data and test data.

Violin maker	year
Catenali (Italy)	1690 ca
A.Stradivari (Italy)	1698
Pietro Guarneri (Italy)	1700 ca
Santo Serafin (Italy)	1700 ca
Graghani (Italy)	1760
Balestrieri (Italy)	1780
Pressenda (Italy)	1838
Fabris (Italy)	1870
Scarpampella (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

2.2 ケプストラム分析と深層学習プログラム

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手順に沿って示す。

録音音源 (WAV, 16Bit PCM 符号あり) を 0.03 sec ずつ切り取り分割保存する。この時、0.01 sec ずつずらして保存を行うため、短い音源データから多くのデータを得ることができる。

得られた短い切り取り音源データを、ケプストラム分析プログラムにかけ、スペクトル包絡を抽出し、その絶対値をラベルの値と共に csv ファイルに格納する。ケプストラムは「波形の短時間振幅スペクトルの対数の逆フーリエ変換」として定義され、スペクトル包絡と微細構造を近似的に分離して抽出することができる分析法である[7]。読み込んだ wav ファイルの波形の標本値系列に時間窓をかけ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。この際、設定したケプストラム次数の値が大きい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 開放弦の音による楽器識別

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。つまり、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る。ケプストラム次数 C (Cepstrum Coefficients) 毎の認識率を計算した結果である。学習用データ数は約 14000, 評価用データ数はおよそ 700 である。ケプストラム次数が 60 あたりで約 95% の精度が得られた。開放弦では楽曲に比べ音の波形が準定常であるため精度が出ているものと思われる。

また, Fig.2 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.2-(a), (b)), $C=20$ では程よくスペクトルの概形をなぞっているが $C=80$ では概形というよりスペクトルのピークをなぞるようになる。よって、一般的にヴァイオリンの楽曲ではさらに 1 オクターブ高い音高もごく普通に譜面に書かれるので, C が大きいと音高の高い時に適切な包絡線が得られなくなる問題が生じる。また G 線でも同様の傾向であった。

以上より、第一段階として楽器の開放弦での識別ができていることが確認できた。

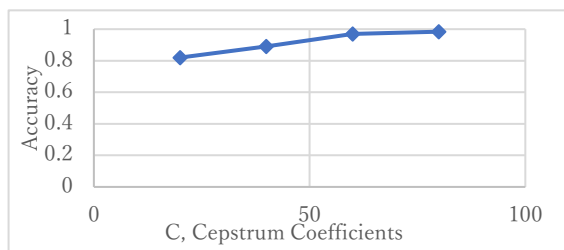


Fig.1 Total accuracy of violin recognition (maker & open strings)

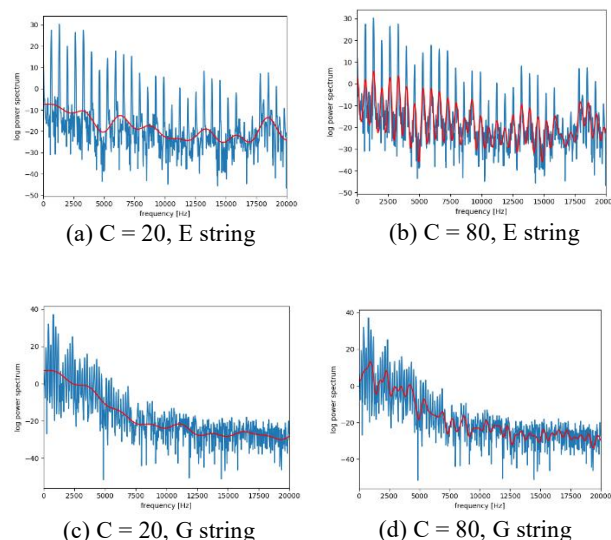


Fig.2 Spectral envelopes with difference in Cepstrum Coefficients, C . (Open E and G string, Fagnola)

3.2 楽曲による楽器識別

演奏家に強弱やビブラートなどの表現を付けて楽曲を演奏してもらった音源による学習用データセットを作成した。開放弦に比べ、音高の変化、ビブラートによる揺らぎ、音量や弦の違い等によりスペクトラムは大きく変動するので、データとしてのバリエーションは多くなる。楽曲「タイスの瞑想曲」のうち最初のテーマの部分で 40 秒ほどの音源である。音域は、下は G 線の A3 から上は E 線の F#6 である。1 つのヴァイオリンにつきこのひと区切りの曲を演奏した録音データに 1 つのラベルを振っている。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

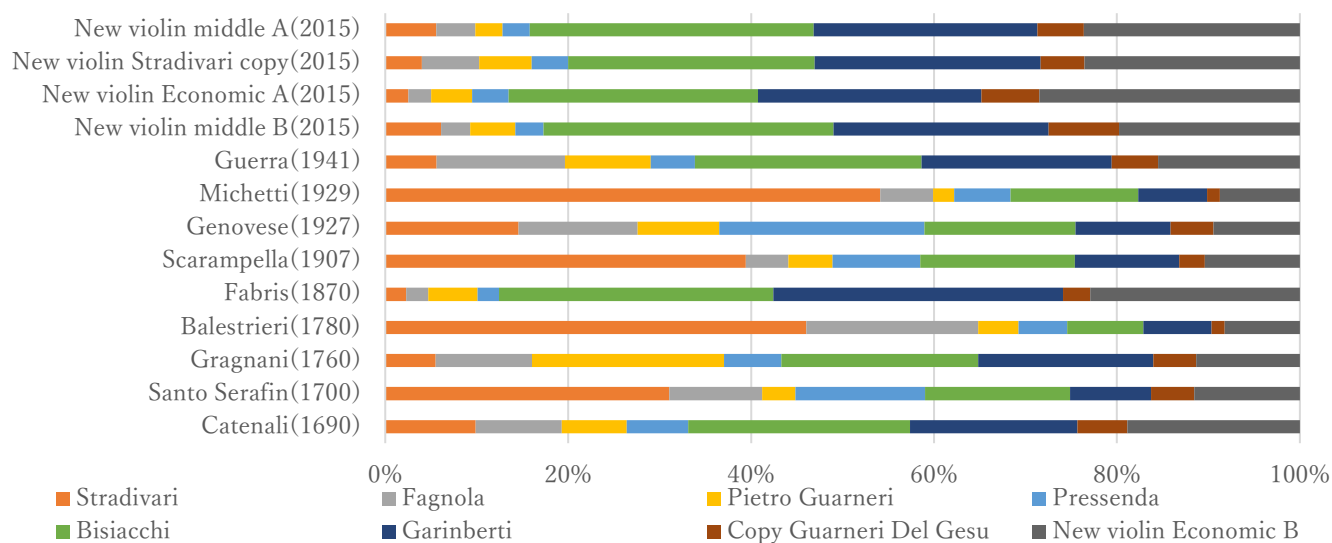


Fig.3 An application of our deep learning program. Percentages of timbre similarity against 8 violins (Stradivari, Fagnola ,,,). The data set is the performance of a piece of music (Meditation de Thais). For example, this chart shows new violins are similar to modern violins such as Bisiacchi or Garinberti, and Michetti is similar to Stradivari at approximately 54%.

約 70000, 評価用データセットが約 3500 である。ケプストラム次数については程よく包絡線が得られる $C=60$ で行った。

結果は, 楽器の学習順序などを変えて行ったが, Accuracy はおよそ 54% から 59% の間であった。やはり, 開放弦に比べ非定常性が大きいと識別の精度は落ちている。スペクトル包絡の計算手法の検討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。

3.3 未知のヴァイオリンの音色傾向分析としてアプリケーション

このヴァイオリン製作者の識別のもう一つのアプリケーション例として, 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, どのような音色の特徴があるかを見積もることができる。つまり, 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アルネリに近いのかの数値的にわかり, 楽器の選択のひとつの指標になることである。

Fig.3 に例を示すが, 8 挺のヴァイオリンを学習用データセットとし, 残りの 13 挺のヴァイオリンを評価に使った場合である。

これは, 横軸に含まれる 8 挺の時代の違うヴァイオリンで学習し, 評価データとして縦軸の 13 挺が, 学習した 8 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。例えば, 上にある 4 つの New violin はモダン・ヴァイオリンの Bisiacchi や Garinberti に合わせて 50% くらい近い音色だといえ, Michetti や Scaranpella は Stradivari にそれぞれ 54%, 39% ほど音色が似ているということを示している。

しかし, 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, おそらく各製作者に対し複数の楽器を (何十挺) 学習させないと正しい判定にはつながらないことは言うまでもない。本研究では今のところ方法論の提案であり, 妥当性は今後の課題である。

4. まとめ

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, 深層学習を用いた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, スペクトル包絡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), 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., Zanoni, M., Antonacci, F., & 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., & 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. *Savart Journal*, 1(3).
- [6] 宮澤芽衣, 横山真男フォルマントに着目したオールドヴァイオリンの音色解析, 日本音響学会 2018 年春季研究発表会予稿集.
- [7] 古井貞熙 (1992) 『音響・音声工学』(電子・情報工学入門シリーズ 2) 近代科学社