Step-QI スクール Step Qi School

楽器の音色をAIで聞き分ける

山田・大寺・北研究室＋学生実験室　倉井 敬史（3年），後藤 悠希（2年）

TOHOKU UNIVERSITY

背景と目的

ガルネリウスやストラディヴァリウスといったバイオリンの名器は、安いバイオリンの音色とどう違うのであろう?素人がその音色の違いを聞き分けることは難しいが、専門家にはそれができる。そこで今回の研修では、バイオリンの音色の違いを科学的に解明し、コンピュータでも聞き分けられるプログラムを開発することを目的とする。

方法としてはまず、音色の異なる2種類のバイオリンで演奏した楽曲を音楽データとしてパソコンに取り込み、機械学習させる。次に、同じバイオリンで演奏した別の楽曲の音楽データを取り込み、どちらのバイオリンで演奏したものかを判別させる。正しく判別できれば、AIが音色の違いを聞き分けていると言える。

楽器の音色の違いとは?

楽器の音色の違いは、基音に対して倍音成分が含まれる割合の違いではないかと考える。情緒的で音楽性豊かな音色を奏でるバイオリンの名器には、含まれる倍音成分に何か違いがあるのではないか?もしそうであるなら、音楽データを周波数領域で解析すれば、その違いが分かるのではないかと考えた。

音楽データの扱い

楽器の音色を周波数領域で解析する手法としては、高速フーリエ変換（FFT）を用いる。音楽データを、20Hz ~ 20kHz の可聴周波数に対して 44.1kHzのサンプリング周波数によりWAVファイルとして取り込むと、時系列での離散データ $h(t)$ が得られる。これを式(1)によって離散フーリエ変換（DFT）することにより、周波数領域の離散データ $H(f)$ が得られる。ここで、N は標本点の数（任意の自然数）である。N の値を決めるにあたり特に注意した点としては、今回聞き分けを行う楽曲(情熱大陸)の音域が 440 ~ 2,000[Hz]であり、音程の間隔が一番狭いところでは約26[Hz]であるので、N=1,500 に設定することによってDFTの周波数分解能が約29[Hz]となり、概ね全ての音程の違いを聞き分けられるようになるので、そのように設定した。(図2)

ニューラルネットワークによる音楽データの学習

音楽データの解析には、パソコンレベルでも短時間で機械学習できるよう、図3に示す様に中間層が一層のみのニューラルネットワークを用いた。処理にかかる時間の関係上、入力層と中間層のノード数を 64としたが、1,500個の標本データを 64個に圧縮する方法として、各標本点の値を、その前後の 23個の値の平均値（1500/64≒23）とした。出力層に関しては、今回は 2台の楽器の聞き分けを行うので 2出力とした。ニューラルネットワークの学習には誤差逆伝播法(Backpropagation)を用いた。これは予め出力層の値に正解を教師データとして与えておき、様々な入力に対して正しい出力が得られるように、各層での重みの値を調整しながら学習していくアルゴリズムである。

実験結果

高価なバイオリンと安価なバイオリンとを用いてそれぞれ演奏した情熱大陸の楽曲(約1分間)の前半部分の約34秒間を教師データとして取り込み、その部分を更に約34msの長さに切り出しながら1000回ニューラルネットワークに学習させた後、次に曲の後半部分を約7秒ごとに分けた検証用データを取り込み、その部分をさらに約34ms の長さに切り出しながら150回比較検証させた。34msという非常に短い長さに切り出すことにより、ポリフォニックな楽曲が単音ごとに分解され、各音の基音と倍音との関係が明確になると考える。比較検証結果を表1のデータ6~9に示す。50%を上回れば正しく識別されたということであり、50%を下回れば誤って識別されたことになり、全ての検証で55%を上回っており、正しく識別されていると言える。また、教師データと同一パートで検証した場合は当然ではあるが、60%以上の高い識別率が得られた（表1のデータ1~5）。上記の比較は、異なる楽器で演奏した同一パート同士の比較であるが、異なるパート間での比較においても検証を行ってみたが、全ての比較検証において55%以上の識別率が得られており、そのような比較においても楽器の違いを識別できていた。

考察とまとめ

楽器の音色の違いは何によるものなのか、また、コンピュータが機械学習によってそれを識別できるものかどうかを確かめる目的で本実験を行った。その結果、楽器の音色の違いが倍音成分の割合の違いにあるものとの仮説のもと、音楽データをFFTを用いて周波数領域で解析し、機械学習をさせることにより、見事に音色の違いを識別することに成功した。

今回の実験では、学習に用いた楽器と検証に用いた楽器は同じものであったが、今後は別の楽器を用いても識別できるのかどうか実験してみたい。また、複数の高価な楽器と安価な楽器とを周波数解析し、高価な楽器と安価な楽器それぞれに共通する特徴を探ることで、楽器音色の違いを決定づけている要素について科学的に解明してみたい。

今回用いたニューラルネットワークは中間層が一層のみ簡単なものであり、多層ニューラルネットワークを用いたディープラーニングに比べて遥かに少ない演算処理で学習を行える。このような簡単な学習モデルでも、人間でも難しい楽器の音色の違いを識別できたことに正直驚いたとともに、機械学習の可能性の大きさを実感した研修であった。

参考文献）
インターフェース　2016年11月号　CQ出版

ストラディヴァリウス　ガルネリウス　1万円のバイオリン

図1　バイオリンの外観の違い

$$H(f)=\sum_{t=0}^{N-1} h(t)\,e^{-i\frac{2\pi f t}{N}}$$

式(1)　離散フーリエ変換(DFT)

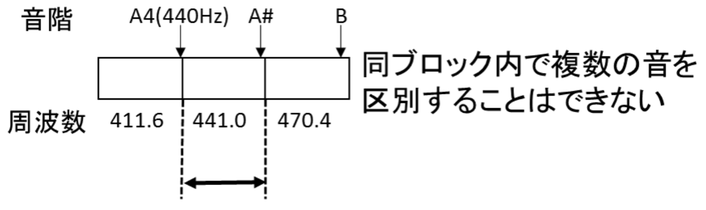

図2　音程の違いを聞き分けるために必要なDFTの周波数分解能

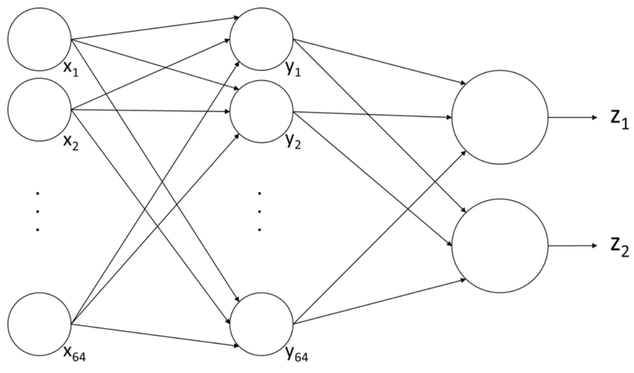

図3　音楽データの解析に用いたニューラルネットワークモデル

表1　音楽データによる検証結果

検証用データと演奏時間	識別率
データ6　[0:30~0:36]	55%
データ7　[0:37~0:44]	58%
データ8　[0:44~0:52]	61%
データ9　[0:52~1:10]	61%
データ1　[0:00~0:06]	61%
データ2　[0:06~0:12]	60%
データ3　[0:12~0:18]	61%
データ4　[0:18~0:23]	62%
データ5　[0:23~0:30]	60%