Vol.2017-MUS-114 No.11
2017/2/27

ニューラルネットワークによる楽器の音色の識別

山田雅之 守田 了

山口大学大学院創成科学研究科

755-8611 宇部市常盤台 2-16-1

本研究では、楽器の音の識別手法を提案する。楽器音の識別について、これまでには関数化された特徴量の近似による楽器音の識別方法が提案されている。それに対して本研究では、楽器音の特徴として倍音成分のパワーを抽出し、これを入力としたニューラルネットワークによる学習を用いた識別方法を提案する。また、音高の種類に対しそれぞれニューラルネットワークを用意することで、音高の変化にのる音色の違いに対応する。この方法は音が人間の耳で電気信号化され脳へ伝わるという流れをモデルにしている。特徴量として倍音構成のみを用い、ニューラルネットワークの学習を用いることでデータ入力から識別までの作業の効率化が図れ、データを用意すれば様々な楽器に対して学習を行い識別することができるという点で有効である。まず楽器の音について FFT をかけ、周波数特性から基本周波数の検出を行い基本周波数とその倍音のスペクトルを抽出する。抽出した値の対数化、量子化を行い、その音高に対応した入力データとしてバックプロパゲーションによる学習を行う．12 種類のオーケストラ楽器による演奏を自ら録音したものと生音のサンプラーによるものを用い $C5$ 音階から $C2$ 音階までの音色を識別する実験を行い有効性を示す。

キーワード 楽器同定、バックプロパゲーション、倍音

Determining Instruments from Sound Based on Neural Network

Masayuki YAMADA Satoru MORITA

The Graduate School of Science and Engineering,Yamaguchi University

2-16-1 Tokiwadai, Ube, 755-8611, Japan

Methods of instrument sounds recognition have been suggested.But consider from various side,unverified helpful methods are exist.Method of recognition instrument sounds by neural network using harmonics this study suggest has not been unverified yet.Flow of this experiment is identify interval from input data including single tone,studying by neural network prepared for each interval difference.This method similar to human's recognition of sound and simple and sure method. Experimental object are tones of 12 instruments in this study.Carrying out frequency analysis for there tones and extract spectrums of fundamental frequency and harmonics. There extracted spectrums were taken logarithm,quantized and input to neural network.As input data of neural network,using 1 10 overtones and studying by back propagation.

keyword: instrument identification, backpropagation, harmonics

1 はじめに

本研究では、ニューラルネットワークを用いた単音の楽器音の識別手法を提案する。楽器音の識別は一般に熟練したオーケストラ奏者でも困難である [1]。これは音の高さや楽器の固体差に応じて様々に音色が変化するためである。テンプレートフィルタリングにより楽器の固体差を、位相トラッキングにより音の高さの揺らぎを吸収することで音色変化の問題に対応して同定する手法が提案されている [2][3]。また音高による音色変化を基本周波数の関数として近似して対応しベイズ決定則を用いて同定する手法が提案されている [4]。

それに対して音色の特徴を楽器音に含まれる倍音成分のパワーの構成として抽出し、ニューラルネットワークによる学習を行う。任意の個数の中間層を入力と出力の間に設けたフィードファワード階層型ニューラルネットワークであるバックプロパゲーションを用いる [5]. 入力に対応する出力のパターンが一致するように各ニューロン間の結合荷重を修正する学習法である。楽器音の固体差は学習に用いる楽器を増やすことで吸収し、聞き分けの難しい聞き取りにくい楽器はニューラルネットワークの構成のうちの隠れ層の階層を増やしたり隠れ層のニューロンの数を変更することで学習性能は向上する。学習時に学習可能であった場合、未知の楽器であっても同種の楽器を学習していると識別可能である。楽器が同定できなくても、あらたにその楽器を学習に加えることで識別が可能になる。

2. では基本周波数ことの倍音成分の楽器ごとの違いを分析する。ニューラルネットワークは大きく2 つに分られる。前半は 2. で述べる基本周波数と倍音成分を抽出する部分である後半は 3. で述べるバックプロパゲーションニューラルネットワークによる楽器音を同定する部分である。4. では実際にオーケストラで使用される 12 楽器について $C5$ 音階から $C2$ 音階までを半音階で音色を識別する実験を自ら録音したものと生音のサンプラーによるもの [6] に対して行い有効性を示す。

2 楽器の音色

2.1 基本周波数と倍音

声や楽器の音などの持つ固有の音質のことを音色という。音色の違いには「倍音」が関係している。コンピュータによる人工音以外の音には基本周波数のほかに様々な周波数のが含まれている。このように音は様々な周波数の音が合成されてひとつの音となり、各周波数の音の含まれ方によって「音色」が生まれる。ヴァイオリンなどの楽器の音にたいして周波数解析をして特性を見ると、数倍音までの成分のパワーが顕著に大きくなっている。本研究ではこの倍音のパワーの大小を楽器の特徴としてニューラルネットワークによる音色の識別を行う。

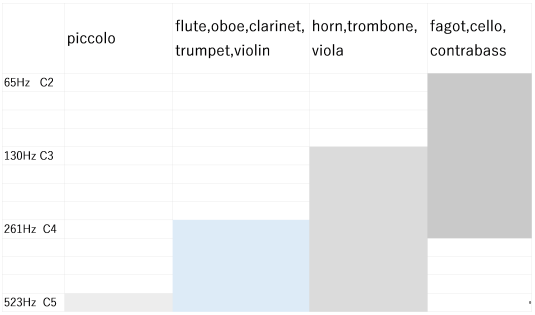


図 1: 楽器の種類と音程
Fig.1:Variation of instruments and pitch extent

2.2 周波数解析

Akai ewi usb garritan の楽器音のサンプル [6] に対して周波数解析を行った。楽器の種類と音程を図 1 に示す。楽器を複数用意し、音量の強弱をかえて録音した音を用いた。

FFT の点数を 32768(2^{15}) として周波数解析を行う。FFT 後の iHz のスペクトル成分 (power[i]) は実部を real[i]、虚部を img[i] として

$$power[i] = (real[i])^2 + (img[i])^2 \qquad (1)$$

として求める。
その結果を2 次元グラフ、3 次元グラフとして示す。図 2,3,4 はそれぞれピッコロ、トランペット、ヴァイオリンの 440Hz の音における周波数特性を表したものである。

図 5,6,7 はそれぞれピッコロ、トランペット、ヴァイオリンにおける、音階の変化に伴う周波数特性の変化を色の違いで表したものである。図 8,9,10 はそれぞれ 130Hz,261Hz,523Hz の音階における、楽器の変化に伴う周波数特性の変化を色の違いで表したものである。これは音の高さや楽器の固体差に応じて様々に音色が変化し、それに応じて倍音成分も変化していることがわかる。このことが一般に熟練したオーケストラ奏者でも楽器音の識別が困難な要因と考えられる、実際に音色を識別する機械を作成することを困難にする要因である。

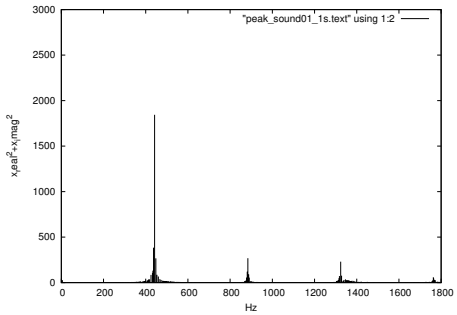


図 2: ピッコロ (440Hz) の周波数特性
Fig.2:Frequency characteristic of piccolo

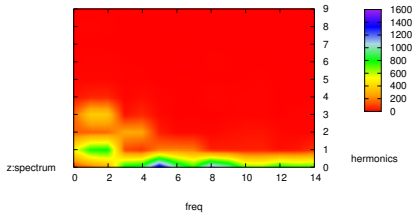


図 5: ピッコロにおける2 オクターブ間C メジャースケールの周波数特性
Fig.5:Frequency characteristic of piccolo in 2octave Cmajor scale

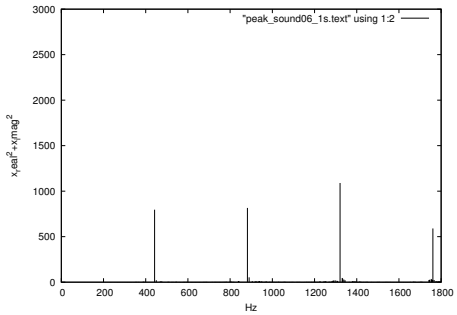


図 3: トランペット (440Hz) の周波数特性
Fig.3:Frequency characteristic of trumpet

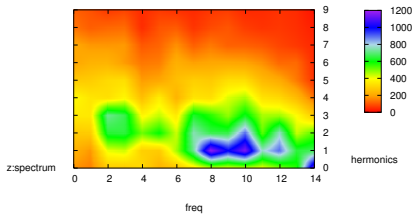


図 6: トランペットにおける2 オクターブ間C メジャースケールの周波数特性
Fig.6:Frequency characteristic of trumpet in 2octave Cmajor scale

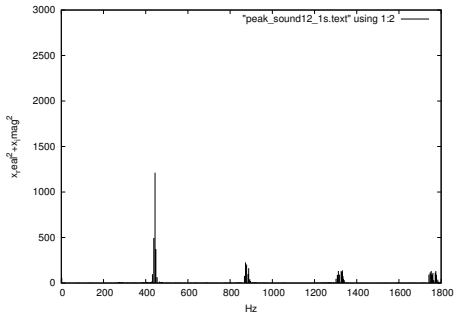


図 4: ヴァイオリン (440Hz) の周波数特性
Fig.4:Frequency characteristic of trumpet

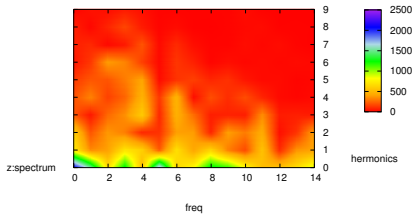


図 7: ヴァイオリンにおける2 オクターブ間C メジャースケールの周波数特性
Fig.7:frequency characteristic of violin in 2octave Cmajor scale

Vol.2017-MUS-114 No.11
2017/2/27

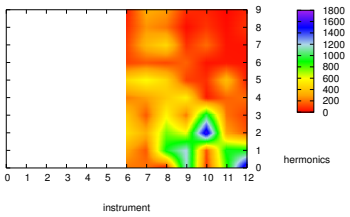


図 8: 13 種類の楽器の 130Hz における周波数特性
Fig.8:Frequency characteristic of 13 instruments in 130Hz

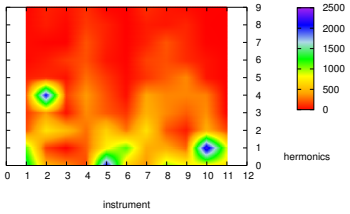


図 9: 13 種類の楽器の 261Hz における周波数特性
Fi.g9:Frequency characteristic of 13 instruments in 261Hz

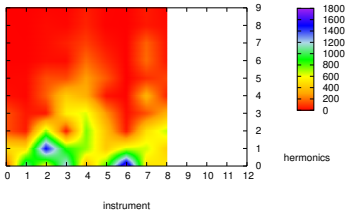


図 10: 13 種類の楽器の 523Hz における周波数特性
Fig.10:Frequency characteristic of 13 instruments in 523Hz

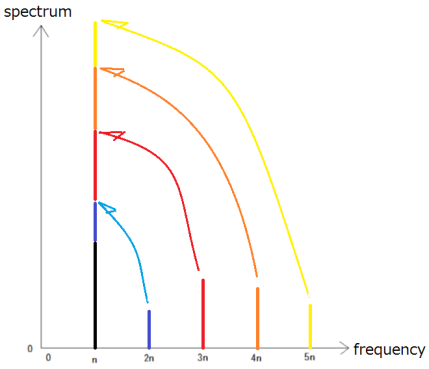


図 11: 基音検出のための倍音の足し 合わせ計算
Fig.11:Detection of fundamental frequency by adding harmonics

2.3 基本周波数、倍音成分の抽出

FFT した後基本周波数と 倍音の成分のスペクト ルの抽出を行う。すべての周波数に対してその 2〜10 倍音の成分のパワーを 加えていく。図 11 は基音検出のための倍音の足し 合わせ計算を 模式的に示している。例えば周波数 iHz に対してこれを 行うとき、iHz の成分のパワーを i、2*iHz を 2i… 10*iHz を 10i と 対応づけ

$$i = i + 2i + 3i + \ldots + 10i \tag{2}$$

のように更新する。楽器の音は倍音成分のパワーが顕著に高く なっているため、その値が大きいものが基本周波数である。足し 合わされた値のリ スト を 大きい順にソートし、最も 大きい値をとった周波数を基本周波数とする。ただし、基本周波数の 1/2 の周波数が検出される場合があったためソートしたときに上位2 つの値の差がわずかな場合、小さい値の方を 基本周波数とする。基本周波数を 推定することができれば、その倍数である倍音の周波数が確定する。また、倍音のスペクト ルを 加える際に誤差を 考慮して、その周辺のピークを 取るために倍数± α の範囲のスペクト ルの最大値を 加える。同じく、基本周波数から 倍音の周波数を 決定する際にも 基本周波数の倍数の± α の範囲での最大値を 倍音のスペクト ルとする。本研究では α =3 と 設定し、基本周波数を 推定した。表 1 は基本周波数検出の誤差率の結果である。計測値 Me 理想値 Id とすると、誤差率 Er の結果は以下の計算式で計算する。

$$Er = ((Me - Id)/Id) \times 100 \tag{3}$$

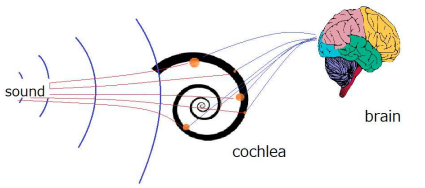
表 1: 基本周波数検出の結果
Table1:Result of fundamental frequency detection

instrument	error(%)
piccolo	0.6423
flute	0.4200
oboe	0.2511
clarinet	0.1702
fagot	0.3346
trumpet	0.1209
horn	0.3900
trombone	0.2308
violin	0.2811
viola	1.1838
cello	0.4635
contrabass	0.7304
all	0.4348

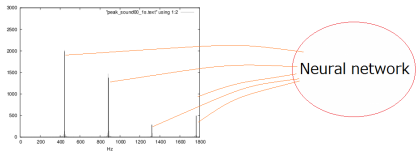
3 ニューラルネットワークによる楽器の音色の識別

人間は周波数が 20Hz から 2 万 Hz の音を聴きとることが出来る。音の波が耳の中に入ると、波が鼓膜を振動させる。この振動は、耳小骨というところで増幅され、その後、奥の蝸牛という場所に送られる。蝸牛は入り口の方は高い周波数の音に反応しやすく、奥に行くほどより低い周波数の音に反応しやすくなっている。この蝸牛がたくさんの神経とつながっているため、どの部分が反応するかによって、異なった神経が活動する。つまり、音の波を周波数ごとに識別して、電気の信号に変換し脳へ送る。その信号は脳幹と呼ばれる場所を経由し、視床、聴覚野の順に送られ音の高さや音色の判別が行われる。蝸牛が様々な周波数の音を入力として反応して脳へ情報を送るように、本研究のニューラルネットワークは 1〜10 倍音のスペクトルを入力として学習を行う。

図 12(a) は人の音色の認識モデルであり、図 12(b) は提案する認識モデルである。耳の蝸牛から周波数スペクトルを受取、音色を知覚している。本研究の認識モデルでも同様に、周波数スペクトルを受取り、そこから倍音と基音を抽出し、音色を知覚するために基音ごとに得られた倍音のスペクトルをもとに音色を知覚している。図 13 は提案する処理の流れを表している。処理の流れは大きく 2 つの部分に分られる。前の節でのべた音声信号から基本周波数と倍音成分を抽出する部分と倍音成分からバックプロパゲーションニューラルネットワーク [5]



(a)



(b)

図 12: (a) 人の音色の認識モデル (b) 提案する認識モデル
Fig.12:Model of tone recognition

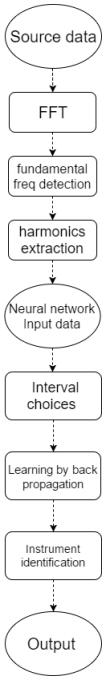


図 13: 処理の流れ
Fig. 13:Flowchart.

を用いて各楽器の音色を学習し、識別する部分である。任意の個数の中間層を入力と出力の間に設けたフィードファワード階層型ニューラルネットワークであるバックプロパゲーションを用いる. 本手法では倍音のスペクトルの入力に対応する出力が対応する楽器と一致するように各ニューロン間の結合荷重を修正する。図 14 は提案するニューラルネットワークの構造の全体である。音の信号からスペクトルを受取、基音と倍音を抽出し、限られた倍音情報からニューラルネットワークで音色を抽出する構造を示している。これにより 12 楽器について複数の楽器が音を出せる範囲を考慮して $C5$ の 523Hz から $C2$ の 65Hz までを半音階で音色を識別する実験を行った。

4　楽器の識別実験

学習および識別に使用するデータには、山口大学医・工学部管弦楽団のメンバーによる演奏の録音を使用する。楽器の種類、音程については先の表 1 で示すものと同様で、各音の長さは2 秒間とする。また、1 種類の楽器について2 つの楽器を用意し奏法はテヌート、スタッカートの2 種類、強弱はメゾピアノとフォルテの2 種類で演奏したものを録音した。(計8 種類) データの形式は WAVE 形式でサンプリング周波数を 44.1kHz、ビット解像度を 24bit とする。

抽出した倍音の中から 10 倍音までのスペクトルを使用し、対数 (log2) をとる。その後 4 ビットに量子化したものをニューラルネットワークの入力データとする。ニューロン数は入力層は 4(bit)*10(倍音) とし隠れ層は入力層と同じ、出力層は 12(楽器) と設定し、バックプロパゲーションにより学習を行う。学習時に学習できない場合は隠れ層の階層を増やしたり数を調整することで学習可能になる。また学習は図 2,3,4, 図 5,6,7, 図 8,9,10, の分析からわかるように音階に応じて連続的に倍音成分は変化するために学習する音階の範囲を小さくして多くのニューラルネットワークを用いて学習するとより確実に楽器を識別できるようになる。出力は認識した楽器に 1 を評価し、それ以外は 0 と評価する。実際にオーケストラに使用される 12 楽器における生音サンンプラー [6] とオーケストラの音を録音したものを用いて、複数の楽器が音を出せる範囲を考慮して $C5$ の 523Hz から $C2$ の 65Hz の範囲で学習可能であることを確認した。

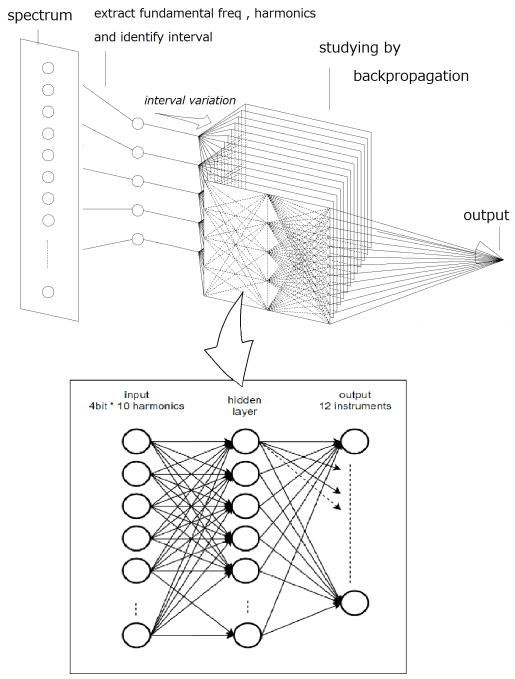


図 14: 提案するニューラルネットワークの構造
Fig.14:Construction of neural network

5　おわりに

本研究では音高毎に基本周波数の量子化された数倍音を入力とするバックプロパゲーションによる学習を行うことで、楽器音の音色の同定ができることを示した。実際に 12 楽器 $C5$ 音階から $C2$ 音階の音色を識別する実験を行い有効性を示した。

謝辞

最後に、録音の際演奏に協力して下さった山口大学医・工学部管弦楽団の太田歩さん、高橋奈歩さん、松葉智子さん、原田佳代子さん、田村萌美さん、秋山晴香さん、三満田翔大さん、細田優海香さん、原田美沙さんに深く感謝致します。

参考文献

[1] Martin, K.D.,"Sound-Source Recognition: A Theory and Computational Model, Ph. D. Thesis, MIT, 1999

[2] 柏野邦夫、村瀬洋、" 適応型混合テンプレートを用いた音源同定 ", 信学論、vol.J81-D-II,no.7,pp.1510-1517,1998

[3] 木下智義、坂井修一、田中英彦、" 周波数成分の重なり適応処理を用いた複数楽器の音源同定処理、信学論、vol.J83-D-II,no.4,pp.1073-1081,2000

[4] 北原鉄朗、後藤真孝、奥野博"音高による音色変化に着目した楽器音の音源同定: F0 依存多次元正規分賦に基づく識別手法", 情報処理学会論文誌, vol. 44, no.10, 2003

[5] D.E.Rumelhart, G. E. Hinton and R. J. Williams, “Learning representation by back-propagation errors, Natrure, vol. 323, pp. 533-536, 1986

[6] "音源: Aria v1.066 Akai ewi usb garritan"