

信号解析とニューラルネットワークを用いた音源同定

2Q-6

佃 卓磨

甲藤 二郎

早稲田大学大学院理工学研究科¹

1 はじめに

自動採譜を行う過程において、楽器の種類と周波数（音高）を認識する音源同定処理が必要である。この音源同定を時間・周波数領域上での信号解析 [1]で行う場合、同一音高だが異なる楽器または同一楽器だが異なる音高として判別される場合が多く、これらが認識率の低下につながっている。

一方周波数スペクトルを Neural Network の入力とする音源同定では、少ない情報量でシンプルなネットワークでも音源同定が有効であると報告されている [2]。本研究では Neural Network は単音で学習させても和音入力に対して音源候補を一つ挙げられることを確認した。この Neural Network の特性を時間・周波数領域上での信号解析に組み込み、音源同定の認識率向上を示す。

2 従来手法

2.1 時間・周波数領域上での信号解析

N 個の音源テンプレート T のうち振幅 w 、位相 p の 2 個の音源の和

$$S(n) = w_i T_i(n + p_i) + w_j T_j(n + p_j) \quad (0 \leq i, j \leq N - 1) \quad (1)$$

で与えられる入力があるとする。

ここで Wavelet 変換した入力とテンプレート $F = W(T)$ の内積から

$$\begin{cases} \langle F_i, W(S) \rangle = w_i |F_i|^2 + w_j \langle F_i, F_j \rangle \\ \langle F_j, W(S) \rangle = w_i \langle F_i, F_j \rangle + w_j |F_j|^2 \end{cases} \quad (2)$$

の連立方程式の解から w を求める。また入力とテンプレートとの位相差 p は相互相関によって求める。これら w と p が求められれば音源分離が可能である。

2.2 ニューラルネットワーク

音源テンプレート T の周波数分布を入力として単音ごとに誤差逆伝搬で学習を行う。本提案方式では、周波数分布は Wavelet 変換を用いて、そのパワー値を Neural Network の入力とする。

3 提案手法

3.1 アルゴリズム

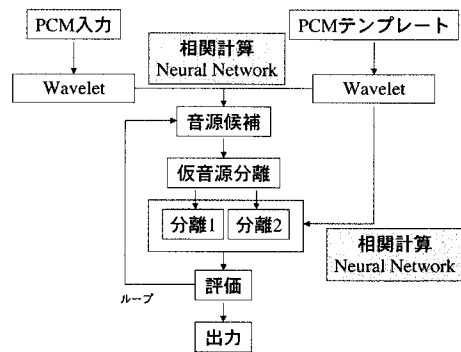


図 1: フローチャート

図 1 に提案手法のフローチャートを示す。まず時間信号の PCM 入力と PCM テンプレートを長い時間窓で Wavelet 変換する。Wavelet 関数は複素関数の代表的な Gabor 関数とし、複素関数は周波数領域上に変換しても位相情報が得られるからである。

次に音源候補を挙げる。和音入力とテンプレートの相関計算を行い、その重み係数の高いものを音源候補とし 1 組 2 音源の候補を得る。またあらかじめ単音テンプレートの Wavelet の周波数分布を Neural Network の誤差逆伝搬で学習させておく。その学習した Neural Network に和音を入力すると 1 音源の候補を得る。音源候補の選び方は、まず相関計算で得られる音源候補 N_c 組を挙げ、その中から Neural Network が挙げた音源候補を含む組を選択する。その後相関計算の高いものから順次 $N_n (N_n < N_c)$ 個選択していく。従来手法は相関計算のみの音源候補

¹ Sound Source Identification Using Signal Analysis and Neural Network
Takuma Tsukuda, Jiro Katto
Graduate School of Science and Engineering, Waseda University

であったが、提案手法は Neural Network からの音源候補も考慮したものになっている。

次に得られた 2 個の音源候補から、式 (2) により求められる振幅 w と位相 p を用いて時間領域上で減算し仮音源分離を行う。分離された音源を相関計算と Neural Network により音源同定を行う。

4 実験

4.1 実験条件

音源テンプレートはサンプリング CD から MIDI 楽器サンプラーで出力し、44.1[kHz]、16[bit] の 2 秒間のデータとする。テンプレートは 3 楽器計 $N = 25$ 個用意し、Guitar が 110, 123, 131, 147, 165, 175, 196, 220, 440[Hz] の $N_g = 9$ 音源、Flute は 440, 494, 523, 587, 659, 698, 784, 880[Hz] の $N_f = 8$ 音源、Sax は 262, 294, 330, 349, 392, 440, 494, 523[Hz] の $N_s = 8$ 音源とする。

Wavelet 変換は時間軸方向に 8[sample]、周波数軸方向に 32[sample] にダウンサンプリングした周波数変換を行い、Neural Network の入力 は $8 \times 32 = 256$ とする。これは和音入力に対して音源候補を 1 つ挙げるために利用する。

音源候補の数は相関計算で得られる候補数 $N_c = 20$ 組から Neural Network で得られる候補数 $N_n = 5$ 組を選択する。

4.2 Neural Network の特性

N 個の音源テンプレート T の周波数分布の学習を実行する際、 N がある程度小さい場合は学習することが容易である。第 4.1 項の Wavelet 変換の実験条件においては、単音テンプレートの学習を行い、同じ単音テンプレートでの音源同定は正解認識率は 100[%] であった。これは完全に学習が出来たということである。

この学習した Neural Network に対して和音入力を与える。 $N = 25$ 個のテンプレートに対しては ${}_{25}C_2 = 300$ 組ある。そのうち Guitar($N_g = 9$) を含む組み合わせは 180 組あり、Flute($N_f = 8$)、Sax($N_s = 8$) は 164 組である。そしてその楽器ごとに正しく認識できたものの個数を数える。また全体の 300 組の 2 音源のうち、どちらか一方を正しく認識できたものの個数を数える。その認識率結果を表 1 に示す。

ただし、和音 1 は学習させたテンプレートと全く同一の波形を加算平均したものであり、和音 2 は MIDI 楽器サンプラーから和音で出力したものをサンプリングしたものである。

表 1: NN 和音認識結果

	Guitar	Flute	Sax	全体
[組数]	180	164	164	300
和音 1 [組数]	49	42	36	127
[%]	27.2	25.6	22.0	42.3
和音 2 [組数]	20	45	80	145
[%]	11.1	27.4	48.8	48.3

4.3 和音評価実験結果

$N = 25$ 個のテンプレートから ${}_{25}C_2 = 300$ の組み合わせの波形を用意する。各音符の発音時間は 5512[sample] でありこれはテンポ 120 で 16 分音符 (0.125[s]) の長さである。

表 2 に和音評価実験の結果を示す。入力 は和音 2 を用いる。入力音符数は 600 でありその中から正しく認識できたものを正解認識率とする。

表 2: 和音正解認識率結果

	Guitar	Flute	Sax	平均
N.N なし [%]	44.9	94.3	69.8	69.2
N.N あり [%]	51.8	91.7	78.6	73.0

5 まとめ

時間・周波数解析での音源同定では、同一楽器だが異なる音高または同一音高だが異なる楽器として誤認識することに注目した。そして Neural Network は単音で学習させても和音入力に対して音源候補を挙げることができ、音源同定の認識率向上の有効性を示した。

参考文献

- [1] 佃卓磨, 甲藤二郎: "テンプレートマッチングを用いた音源同定に関する一検討", 2001 年電子情報通信学会総合大会 D-14-10
- [2] 村瀬樹太郎, 中西正和: "ニューラルネットワークを用いた複数楽器の音源同定処理", 音楽情報科学 39-18(2001.2.23)