

F-003

メロディの遷移確率を考慮したニューラルネットワークによる音源同定 Sound Source Identification Using Hopfield-Type Neural Networks with Melody Transition Probabilities

久保田 優†
Yuu Kubota

甲藤 二郎‡
Jiro Katto

1. まえがき

複数の音響信号が重畳された信号を入力とし、それに含まれる楽器音を推定する音源同定に関する研究が盛んである。著者らはこれまでに、音源同定の問題を組み合わせ最適化問題と位置づけ、Hopfield 型のニューラルネットワークを用いて最適解を求める方法を試みた。[1]

しかし、ニューラルネットワークを用いた解法においては局所解に陥ってしまうことがしばしば問題となる。

そこで、本稿ではニューラルネットワークの局所解の問題を軽減するために、楽曲のメロディの遷移確率を考慮した補正を行い、音源同定精度の向上を図る方法を提案する。

2. 従来手法と問題点

著者らがこれまでに発表した方法[1]の概要は、あらかじめ音源同定を行う対象の信号に含まれている可能性のある全ての楽器・音高の単音の演奏データを録音し、ウェーブレット変換を施した後に、その周波数スペクトルをテンプレートとして用意しておく。そして、式(1)に従って、テンプレートを組み合わせて入力信号と同等の信号を合成するイメージ、すなわち単音テンプレートの組み合わせ最適化(エネルギー最小化)問題として音源同定を捉えたものである。

$$E = K_1 \cdot \frac{1}{2} \sum_{f=1}^F \left(\sum_{i=1}^N x_i T_{if} - S_f \right)^2 + K_2 \cdot \frac{1}{2} \sum_{i=1}^N \sum_{f=1}^F T_{if}^2 x_i (1 - x_i) \quad (1)$$

上式において、 F はスペクトルの点数、 N は総テンプレート数、 x_i はニューロン i の出力、 T_{if} は i 番目のテンプレートのスペクトルにおける f 番目の成分、 S_f は入力信号のスペクトルにおける f 番目の成分、そして K_1 と K_2 は各項の寄与率である。また後項はネットワークの収束性を上げるための補正項である。式全体として、出力が正しい音の組み合わせを示す時、エネルギー E は最小となる。

この手法は処理が高速で複数音同時認識が可能である反面、テンプレートの類似度が高いもの同士で誤認識を起こすという問題点がある。この結果、「オクターブ誤り」や「楽器誤り」を引き起こす。これらはニューラルネットワークが最適解探索に失敗し、局所解に陥ってしまったものと考えられる。

そこで、本稿では音源同定に音楽的な要素を加味し、音楽的な視点からもアプローチすることで音源同定精度の向上を図る。

3. 提案手法

3.1 提案手法概要

2. で求められたニューラルネットワークによる音源同定結果にメロディの遷移確率を考慮することにより、オクターブ誤りを始めとする音楽的に不自然な誤りを防ぐ(局所解を回避する)ことができると考えられる。合わせて、前処理として、最小二乗法を用いた探索テンプレートの絞込みを行う。処理の流れを図1に示す。

本提案手法において重要となる考え方は、式(1)で定義したように、「ニューラルネットワークの出力が正しい音の組み合わせを示している時にエネルギーが最小になる」ということである。

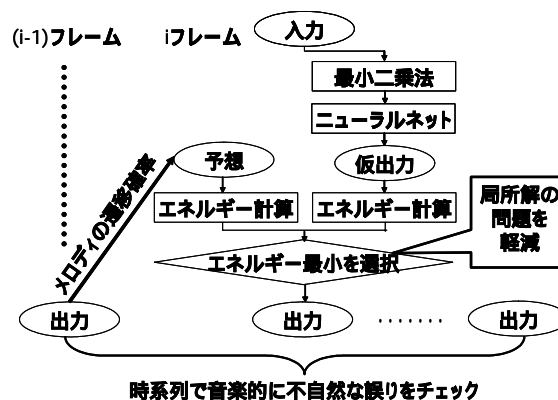


図1. 処理の流れ

3.2 最小二乗法による探索テンプレートの選抜

用意するテンプレートの数が多いほどそれだけ組み合わせの場合の数が増えるので、ニューラルネットワークによる解法は局所解に陥りやすくなる。したがって、ニューラルネットワーク部の前処理として、予め最小二乗法でテンプレートの候補を絞っておけば局所解の問題を軽減できると考えられる。

図1の最小二乗法部では、式(1)の前項が示す入力信号と、予め用意した全てのテンプレートの重ね合わせの誤差の二乗和が最小になるように式(2)を用いて x_i を推定し、入力信号に含まれている可能性が高い音に対応するテンプレートを上位複数個抽出する。

\mathbf{x} の最小二乗推定値 $\hat{\mathbf{x}}$ は

$$\hat{\mathbf{x}} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_N]^T = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{S} \quad (2)$$

で与えられる。ここで、 \mathbf{T} は T_{if} をまとめた $(F \times N)$ 行列、 \mathbf{S} は S_f をまとめた F 次元のベクトルで表現したものである。

† 早稲田大学大学院理工学研究科

‡ 早稲田大学理工学部

3.3 メロディの遷移確率の統計結果

音楽には“ありがちな”メロディーラインやコード進行がある。そこで器楽の教科書[4]等に掲載している比較的シンプルなもの 40 曲 (総音数 1686) を対象に統計をとり、メロディがどのように遷移するかという確率を求めた。本稿では、これを「メロディの遷移確率」と呼ぶことにする。その統計結果を図 2 に示す。

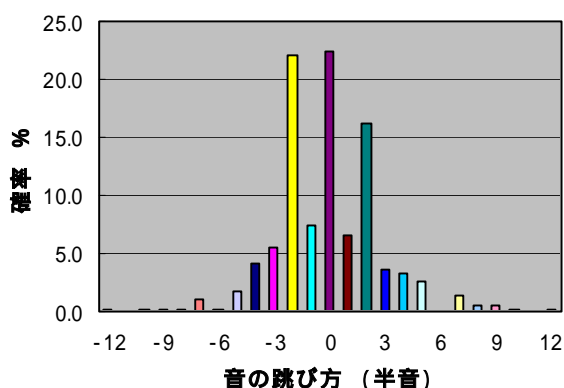


図 2 . メロディの遷移確率

図 2 からわかるように、大きく音高が跳ぶことは少なく、7 割程度の確率で上下全音 1 音以内に遷移する傾向が見られた。

また、本稿ではメロディの遷移確率は 1 つ前の音高のみに依存し、かつ定常な遷移確率を持つとみなすことにする。

3.4 メロディの遷移確率を用いた出力予測

まず、図 1 の右側の流れを見ていくと、現フレーム (i フレーム) において入力信号から最小二乗法を用いて選出したテンプレートによりニューラルネットワークを構築し、2. で示した方法で音源同定を行って、その結果を仮出力として保持する。ここで、本提案手法では図 1 の左側に示したように、1 つ前のフレーム ($i-1$ フレーム) の出力を参照し、メロディの遷移確率を考慮することにより、現フレーム (i フレーム) で出やすい音の組み合わせの候補をいくつか予想する。

そして、ニューラルネットワークによって求められた音の組み合わせのエネルギーと、メロディの遷移確率によって予想された音の組み合わせのエネルギーを式(1)を用いてそれぞれ計算をする。ここで、改めてそれらの中で最もエネルギーの小さいものを選び、その音の組み合わせを新たな出力とする。これにより、ニューラルネットワークを用いた最適解の探索において大きな障害となる局所解の問題を軽減できると考えられる。つまり、メロディの遷移確率は局所解に陥ってしまった場合にそこから脱却するための手段として用いる。

また、上に示した手順で求められた出力を時系列で数フレーム並べてチェックし、音楽的に不自然な出力があればその前後から判断し、必要に応じて訂正するようにした。

4 . 実験

入力には器楽の教科書[4]等に掲載している既存の曲を用いた。曲には単音・2 和音・3 和音のものを用意したが、中には途中で和音数が変わったり、休符が入っている曲も存在する。予め用意する単音テンプレートとして RWC 研究用音楽データベース[5]から以下の 3 楽器に関し、

- (a)ギター : A2 ~ C4# (110 ~ 277[Hz]) 17 音
- (b)ピアノ : A2 ~ C4# (110 ~ 277[Hz]) 17 音
- (c)トランペット : A3 ~ C5# (220 ~ 554[Hz]) 17 音

それぞれ 17 個ずつ、計 51 個の実音源のテンプレートを用意した。これらは実際の演奏音域を考慮するとともに、別楽器で同じ音高のテンプレートが存在するように選択した。また、入力とする曲もこのテンプレートを組み合わせることで作成し、入力信号として音源同定する処理単位に手動で切り出した。

実験の評価方法は[3]を参考に以下の式で認識精度を与えることとする。

認識精度 [%] =

$$\frac{1}{2} \times \left(\frac{\text{出力中の正解数}}{\text{出力された単音数}} + \frac{\text{出力中の正解数}}{\text{正解中の単音数}} \right) \times 100$$

音高と楽器がともに全て一致した場合、認識精度は 100% となる。

表 1 . 認識結果

入力	実験曲数	認識精度 [%]	
		従来手法	提案手法
単音	15 曲 (音数 736)	93.7	93.9
2 和音	12 曲 (音数 1139)	75.0	92.6
3 和音	3 曲 (音数 404)	62.4	79.8

5 . おわりに

本稿では、ニューラルネットワークを用いた音源同定において楽曲のメロディの遷移確率を考慮することで、局所解の問題を軽減する方法を提案し、結果として音源同定精度を向上させることができた。

今後は、振幅の変動やテンプレートに用意していない音にも対応できるように改良していきたいと考えている。

6 . 参考文献

- [1] 久保田優, 甲藤二郎, “Hopfield 型ニューラルネットワークを用いた音源同定”, 電子情報通信学会 2003 年総合大会講演論文集, D-14-22, p.189, 2003
- [2] 柏野邦夫, 村瀬洋, “アンサンブル実演奏の自動アンミキサ”, 信学技報, SP97-104, pp.33-40, 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”, 電子情報通信学会論文誌, Vol.J83-D- , No.4, pp.1073-1081, 2000
- [4] 市川都志春, “中学生の器楽”, 教育芸術社, 1996
- [5] RWC 研究用音楽データベース,
<http://staff.aist.go.jp/m.goto/RWC-MDB/index-j.html>