

スペクトル領域の DP マッチングによる自然楽器演奏の多重音解析*

亀岡弘和 篠田浩一 嵯峨山茂樹 (東大・情報理工)

1 はじめに

音楽演奏における多重音のピッチ推定の手法としては、楕円フィルタを用いた方法 [2]、混合ガウス分布モデルを用いた方法 [3] や、音源ごとに各単音のスペクトルをテンプレートとしてマッチングを行う方法 [4, 5] などがある。これらの研究の中で挙げられる問題の一つは、自然楽器演奏に生じるピッチの揺らぎがピッチ推定に悪影響を及ぼすということである。

本報告では、このような問題に対し、スペクトルをテンプレートとして非線形なマッチングをスペクトル領域の動的計画法 (DP) により行う方法を提案し、自然楽器演奏を用いて検証を行う。

2 スペクトル領域の DP マッチング

2.1 自然楽器のピッチの問題

自然楽器音 (ヴァイオリン、歌唱など) においては、ピッチが平均律から外れることがある。図 1 の (a), (b) 上段は、平均律から外れたピッチの単音スペクトルを表す。このようにピッチにずれがある単音に対し、単純に周波数ごとのスペクトル同士を対応させてマッチングを行う (以後単純マッチングと呼ぶ) 場合は平均律のピッチのスペクトルテンプレートのみでは不十分であり、スペクトルテンプレートを周波数軸方向に伸縮させ、ピッチのずれに対応する新たなスペクトルテンプレートを用意する必要がある。また、和音のスペクトルテンプレートは構成音のスペクトルテンプレートを加算させて生成するため、図 1(c) 上段のように構成音のピッチが独立にずれをもつ和音のスペクトルの場合、ピッチのずれに対応するには構成音数のべき乗に比例する個数のスペクトルテンプレートが必要となる。従って、同時発音数の多い音響信号を対象を発展させる場合、現実的とはいえない。

2.2 周波数領域の DP マッチングの適用

以上のような問題に対し、入力音響信号とテンプレートとのスペクトルのマッチングに DP マッチングを適用する方法を提案する。入力信号のスペクトル、スペクトルテンプレートそれぞれの系列 A と B を、

$$A = a_1, a_2, \dots, a_i, \dots, a_N \quad (1)$$

$$B = b_1, b_2, \dots, b_j, \dots, b_N \quad (2)$$

とする。 $d(i, j)$ を a_i と b_j のユークリッド距離とすると、二つのスペクトル系列の累積距離 $g(i, j)$ を次式により求める。

$$g(i, j) = \min \left\{ \begin{array}{l} g(i-1, j) + d(i, j) \\ g(i-1, j-1) + 2d(i, j) \\ g(i, j-1) + d(i, j) \end{array} \right\} \quad (3)$$

これを、 $(i, j) = (1, 1)$ から $(i, j) = (N, N)$ まで順次求めていき、 A, B 両パターン間の最小累積距離 $g(N, N)$ を得る。ここで、平均律で隣り合うピッチ間の中間までの範囲内となる傾斜制限、

$$2^{-\frac{1}{24}} \cdot i < j < 2^{\frac{1}{24}} \cdot i \quad (4)$$

*“Multi-Pitch Analysis of Natural Instruments Performances using DP Matching in Spectrum Domain” by Hirokazu KAMEOKA, Koichi SHINODA and Shigeki SAGAYAMA (Graduate School of Information Science and Technology, The University of Tokyo).

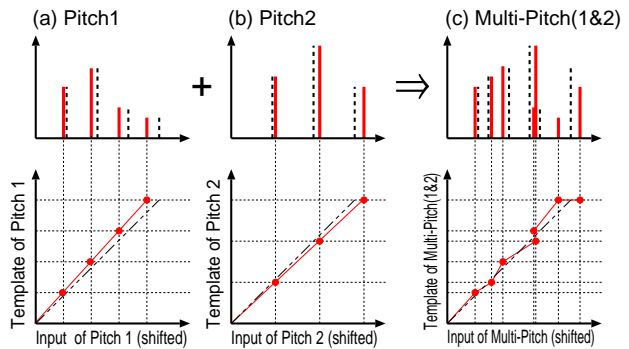


図 1: ピッチに独立にずれをもつ構成音 (a), (b) からなる和音 (c)

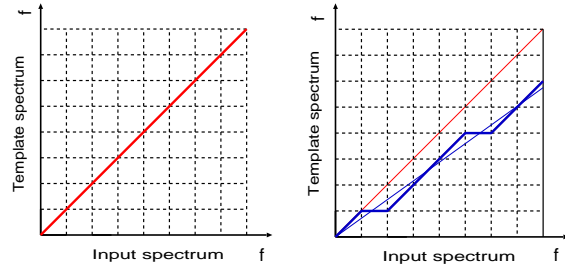


図 2: ピッチにずれがない場合

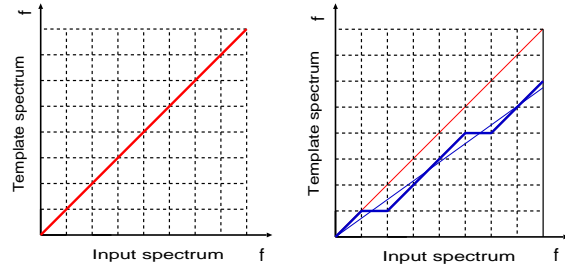


図 3: ピッチにずれがある場合

を設ける。以上の提案方法を用いることで両系列の対応付けを非線形に行うことができ、以下の二つの問題が解決される。

(i) 構成音ピッチの独立なずれの問題

ピッチにずれをもつ単音の入力信号スペクトルとその音階のスペクトルテンプレートの対応関係は、入力ピッチにずれがある場合、図 1(a), (b) 下段のようになる。また、各音階を構成音とする和音スペクトルテンプレートと、二つの構成音のピッチが独立にずれをもつこのような和音のスペクトルとの対応関係は図 1(c) 下段のように非線形になる。

(ii) 離散的スペクトルの問題

スペクトル解析処理では高速フーリエ変換 (FFT) などを用いてスペクトルを離散的に扱うことが多い。図 2 はピッチにまったくずれがない場合の、図 3 はピッチにずれをもつ場合の対応関係を表す。図 3 のようにピッチにずれがある場合、対応関係は格子点上で階段状になり、非線形である。特に低音に関しては、スペクトル成分の間隔は狭くなるが格子点間隔は一定なので、大きい影響が予想される。

3 処理系

非線形なスペクトルマッチングの効果を調べるため、以下のような処理系 (図 4) を構成する。

(i) セグメンテーション

音符長に相当し構成音のピッチが一定である区間を「セグメント」と定義する。FFT の解析区間を以後フレームと呼ぶことにすると、入力スペクトル系列から、音の立ち上がりが見なされるフレームまたは構

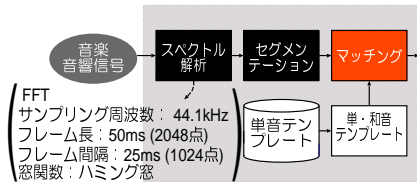


図 4: 処理系の概要図

成音が変化したと見なされるフレームを探し、セグメントの開始フレームとする。また、セグメントの終了時点は次のセグメントの開始時点の手前のフレームとする。

(ii) テンプレート

曲ごとの最低音から最高音までの半音間隔すべてのピッチのスペクトルを単音テンプレートとする。今回は、各楽器音の立ち上がり直後の1フレーム分のスペクトルを目視で選択し、単音テンプレートとした。最大3つまでの単音テンプレートを加算生成したものを和音テンプレートとする。

(iii) マッチング方法

スペクトルテンプレート x と入力信号のフレーム t のスペクトルと DP 累積距離を $D(x, t)$ とすると、

$$x^* = \underset{x}{\operatorname{argmin}} \{ \min_t D(x, t) \}$$

となるテンプレート x^* を推定結果とする。

(iv) 出力結果の補正

楽器音の立ち上がり時や消音時付近で余分なセグメントを設けてしまい、推定結果にノイズを含むことがある。推定ピッチ系列を HMM[9] でモデル化し、Viterbi アルゴリズムによりノイズの除去を行う。

4 評価実験

プロの演奏家による自然楽器（ピアノまたはヴァイオリン）演奏を音響信号を入力とし、各演奏曲のピッチを推定した。演奏の最大同時発音数は3とした。また、DP マッチングと単純マッチングを用いた推定結果を比較した。

表 1 に DP マッチングと単純マッチングそれぞれの方法による各曲の和音の認識率を、表 2 に音高の認識率を示す。また、表 3 に高音域 (C4 以上)、低音域 (B3 以下) それぞれについての各手法の認識率を示す。結果から、演奏された曲・楽器、音域にかかわらず、DP マッチングを用いた方が高い認識率を得ることができた。ヴァイオリンによって演奏された曲、または低音域に関して、特に高い認識率を得た。前者

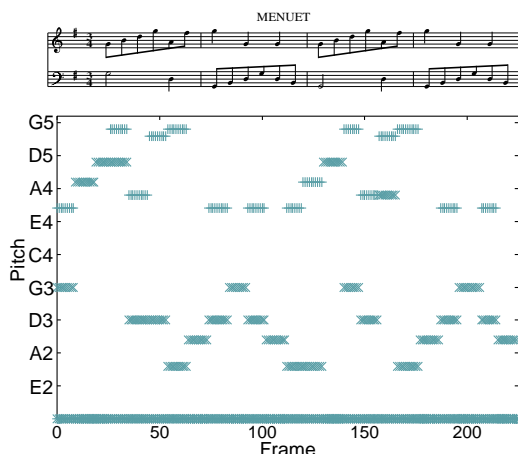


図 5: Bach: "Menuet" の一部の譜面と推定結果

表 1: 和音の認識率

曲名	楽器	和音数	方法 (%)	
			DP	単純
Bach: "Menuet"	Piano	205	81.6	78.6
Mozart: "Menuet"	Piano	58	86.2	79.3
Pachelbel: "Kanon"	Violin	743	71.6	53.4

表 2: 音高の認識率

曲名	音符数	マッチング方法 (%)	
		DP	単純
Bach: "Menuet"	309	92.9	90.0
Mozart: "Menuet"	97	90.7	85.6
Pachelbel: "Kanon"	1638	86.7	78.3

表 3: 高・低音の認識率

音域	マッチング方法 (%)	
	DP	単純
高音 (C4 以上)	91.9	91.0
低音 (B3 以下)	90.1	81.6

はヴァイオリンがピッチの揺らぐことが多い楽器であること、後者は低音であればあるほどスペクトル成分の間隔は狭くなるにもかかわらずマッチング格子点間隔は一定であることで非線形性の影響がより顕著になって現れたことが理由として考えられる。

5 おわりに

本報告では、自然楽器演奏のピッチのずれに対して頑健な和音ピッチ推定法として、スペクトル領域の DP マッチングによるスペクトルマッチング方法を提案した。提案方法は、単純マッチング方法と比較して全体的に優れた結果を得た。ピッチが揺らぐことの多いヴァイオリン、および低音域において特に優位に働くことが分かった。

DP マッチングは効率のよい計算方法であるが、今後ビームサーチなどにより計算効率をさらに向上させたい。また、音色の時間変化をより反映した手法を検討していきたい。

謝辞

本研究の一部は、科学技術振興事業団戦略的基礎研究推進事業 (CREST) (「脳を創る」聴覚脳研究プロジェクト) の支援を受けて行われた。

参考文献

- [1] 白土保: "二重奏音からの基本周波数分離抽出," 日本音響学会誌, Vol. 54, No. 10, pp. 715-719, 1998.
- [2] 三輪多恵子, 田所嘉昭, 斎藤努: "くし形フィルタを利用した採譜のための異楽器音中のピッチ推定," 電子情報通信学会論文誌, Vol. J81-D-II, No. 9, pp. 1965-1974, 1998.
- [3] 後藤真孝: "音楽音響信号を対象としたメロディーとベースの音高推定," 情報処理学会研究報告, 99-MUS-31-16, Vol. 99, No. 68, 1999.
- [4] 中臺一博, 柏野邦夫, 田中英彦: "音楽音響信号を対象とする音源分離システム," 情報処理学会技術研究報告, SIGMUS1-1, pp. 1-8, 1993.
- [5] 小野徹太郎, 斎藤英雄, 小沢慎治: "自動採譜のための GA を用いた混合音推定," 計測自動制御学会論文集, Vol. 33, No. 5, pp. 417-423, 1997.
- [6] 柏野, 木下: "音楽情景分析の処理モデル OPTIMA における和音の認識," 電子情報通信学会論文誌, Vol. J79-D-II, No. 11, pp. 1762-1770, 1996.
- [7] 柏野, 村瀬: "適応型混合テンプレートを用いた音源同定," 電子情報通信学会論文誌, Vol. J81-D-II, No. 7, pp. 1510-1517, 1998.
- [8] 木下, 坂井, 田中: "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-II, No. 4, pp. 1073-1081, 2000.
- [9] 古井貞照: "音声情報処理," 森北出版, 1998.