

周波数領域の DP マッチングによる自然楽器演奏の和音ピッチ推定

亀岡 弘和 篠田 浩一 嵯峨山茂樹

東京大学大学院情報理工学系研究科

〒 113-0033 東京都文京区本郷 7-3-1

E-mail: †{kameoka,k-shino,sagayama}@hil.t.u-tokyo.ac.jp

あらまし 本報告では、自然楽器による和音演奏のピッチ推定を扱う。音楽音響信号の音符長に相当するセグメントの入力スペクトログラムと、楽器音の代表的なスペクトルテンプレートとを周波数領域で DP マッチングを行うことで、ピッチを推定する方法を提案する。セグメントは、入力信号のパワー変化やスペクトル成分の変化に着目して分割して作成する。提案方法と従来の単純なスペクトルマッチング方法と比較し、その有効性についての検証を行った。ピアノ 1 台による演奏データに対する音高の認識率は 92%、和音認識率は 82%であった。ヴァイオリンの 3 重奏データに対する音高の認識率は 87%、和音認識率は 71%であった。また、従来のマッチング方法と比較して、弦楽器または低音についての優位性が示された。

キーワード 自動採譜, 自然楽器演奏, 和音ピッチ推定, スペクトルマッチング, DP マッチング

Multi-Pitch Estimation of Natural Instruments Performances by using DP Matching in Frequency Domain

Hirokazu KAMEOKA, Koichi SHINODA, and Shigeki SAGAYAMA

Graduate School of Information Science and Technology, University of Tokyo

7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033 Japan

E-mail: †{kameoka,k-shino,sagayama}@hil.t.u-tokyo.ac.jp

Abstract This report deals with an pitch estimation method for multi-pitch performances with natural instruments. First, musical audio signal stream is divided into segments. The length of each segment corresponds to that of each note. Secondly, in each segment, template matching using Dynamic Programming is applied between the input spectrogram and every spectrum template which is synthesized by tone templates. Each tone template is a representative spectrum of scale played by an instrument. In the experimental results, the accuracy of pitch estimation and chord estimation were 92% and 82% when using piano performances as musical data. And when using violin ensemble, the accuracy of pitch estimation and chord estimation were 87% and 71%. The accuracy of proposed method was higher than that of a simple template matching method especially at lower pitch or when using string instruments.

Key words music transcription, natural instruments performance, multi-pitch estimation, spectrum template matching, DP matching

1. ま え が き

作曲活動や即興演奏の記録などの音楽活動全般において採譜作業は必要不可欠である。その採譜作業をコンピュータによって行えるようにすることが、自動採譜研究の目的である。

自動採譜の研究はこれまでに数多く試みられている。1990 年頃までは単一楽器による単音による演奏が対象で、近年になって単一楽器または複数楽器による和音の演奏へと対象を拡張してきた [1], [5], [6], [8], [9]。しかし和音は各音の高調波関係にあ

る周波数成分が重なり合うため、一般に和音を対象としたピッチ推定は難しいとされている。例えば、二重奏音から個々の単音の基本周波数分離抽出する方法 [1] では、一つの単音のみに含まれる倍音周波数成分のローカルピークを対象音のスペクトルからいくつか見つける必要があるが、倍音周波数成分が重複してしまうとそれらを見つけることが困難であるとされている。また、倍音構成に着目した方法 [2], [3] では、倍音周波数成分の中で最大のパワーをもつスペクトルからピッチ推定を行うため、同様に推定が困難であるとされている。

そこで、この問題に対するアプローチとして、楕円フィルタを用いた方法 [4] や、音源ごとに各単音のスペクトルをテンプレートとし、マッチングを行う方法 [5], [6] などが提案されている。楕円フィルタを用いた方法では、周波数領域において等間隔に零点を有するフィルタの次数を変化させて非零出力を検出することで和音を構成するピッチを順次判別できる。この方法は、演奏楽器の種類や楽器音の音色の時間変動に依存せずピッチ検出が可能であるという利点をもつ。しかし楽器判別は、楕円フィルタによる分離波形と各楽器の基準波形との相関値により行うため、異種楽器が同一音を演奏した場合の判別が困難である。それに対し、スペクトルのテンプレートマッチングを行う方法では、同時発音数やテンプレート数の増加に伴って計算量が膨大になるという問題はあるものの、異種楽器それぞれのスペクトルを加算生成したテンプレートを用意しておけば、同一音を演奏した場合でも判別は可能である。

自動採譜の分野では、ピッチ推定と楽器判別共に重要な要素の一つである。そこで今後、ピッチ推定だけでなく楽器判別機能への拡張を考慮に入れ、この手法を用いたピッチ推定法について考える。

弦楽器や歌唱ではピブラートや過渡的なピッチ変化や技量的な限界などがあるため、音楽一般を対象として扱うとき、ピッチのずれは避けられない現象である。しかしこの手法は従来、各音階に対応するスペクトルテンプレートしか用いないため、ピッチのずれによるスペクトル変動の影響を厳密に吸収できる手法ではなかった。

以下の章では、その問題の対処として DP マッチングの導入について述べる。DP マッチングは、音声認識の分野で広く用いられているアルゴリズムである [16], [17]。音声認識の手法として、入力音声とテンプレートのパターンマッチングを行う方式がある。一般に人間の発声時間長は同じ話者でも発声のたびに不規則に変動するため、入力音声パターンと標準パターンをそのまま比較することができない。そこで、二つのパターンの時間正規化を行い、両パターンの最適なマッチング系列を求めるのにこのアルゴリズムが用いられる。

本報告では、スペクトルのテンプレートマッチングを行う際のピッチのずれによって生じる問題を考え、その解決策として音声認識の分野の代表的なアルゴリズムである DP マッチングを周波数領域で用いる方法を提案する。また、自然楽器の実演奏を用いて実験を行い、その有効性についての検証を行った結果を述べる。

2. 周波数領域の DP マッチング

2.1 自然楽器のピッチの問題

先述した通り自動採譜の対象として音楽を一般的に扱うとき、ピッチのずれを不可避な現象として考える必要がある。楽器音はその音色を決定する倍音と基本波によって構成され、一般的にこれらは調和関係が保たれる（ピアノ音は高周域のスペクトル成分は非調和関係にある [10]）。従って、単音のピッチが本来の音階からずれる場合、スペクトル成分は周波数軸に線形伸縮する。構成音のピッチが独立にずれをもつ（構成音 C, E, G の

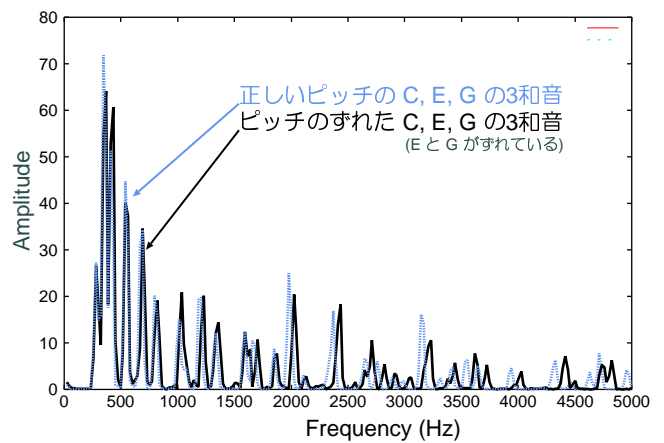


図 1 ヴァイオリンによる和音のスペクトル

うち E, G にずれがある) 和音とピッチにずれがない和音のスペクトルを図 1 に示す。

図 2 の (a), (b) 上段は、本来の音階からずれたピッチの単音スペクトルを表す。このように単音のピッチにずれがある場合、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）上でそれに適合したスペクトルテンプレートを生成するためには、テンプレートを周波数軸に線形伸縮を行う必要がある。また、図 2(c) 上段のようにピッチにずれがある単音からなる和音のスペクトルに適合したスペクトルテンプレートは、各構成音スペクトルの線形伸縮を行ったものの重ね合わせで生成される。従って、そのような和音に対してスペクトルテンプレートを適合させるには、構成音ごとにスペクトルを線形伸縮して組み合わせなければならないため、構成音数のべき乗に比例する計算が必要になる。

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。その理由を以下に述べる。

(i) 構成音ピッチの独立なずれの問題

単音ピッチの入力信号とその音階のスペクトルテンプレートとのスペクトル間の距離が最小となるマッチング経路（以後、最小マッチング経路と呼ぶ）は、入力ピッチにずれがある場合、図

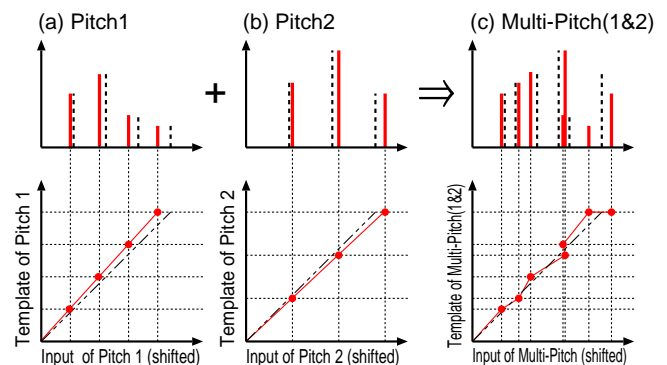


図 2 ピッチに独立にずれをもつ構成音 (a), (b) からなる和音のスペクトル間最小マッチング経路 (c)

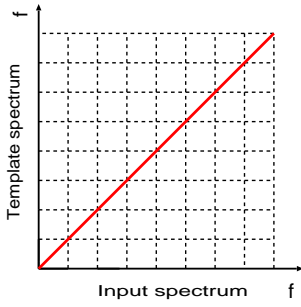


図3 ピッチにずれがない場合

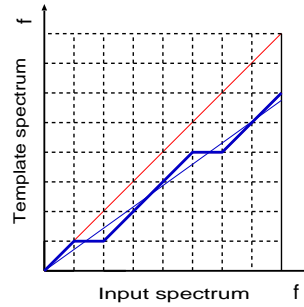


図4 ピッチにずれがある場合

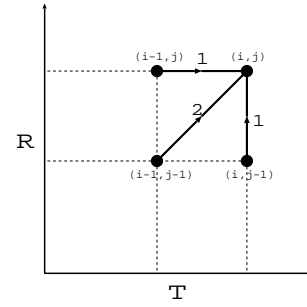


図5 対称形

2(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間での最小マッチング経路は図2(c)下段のように非線形になる。ただし、図2下段は横軸を入力信号のスペクトル、縦軸をテンプレートのスペクトルとし、その最適なマッチング対応経路を示した図である。

(ii) 離散的スペクトルの問題

スペクトル解析処理ではFFT(Fast Fourier Transform)などを用いてスペクトルを離散的に扱うことが多い。下図は離散的な入力スペクトルパターンとテンプレートスペクトルパターンとのマッチングの最適な対応経路を表した図である。図3は、ピッチにまったくずれがない場合の最適な対応経路を表しており、この場合対応経路が直線なので単純マッチングを用いても問題はない。しかし、スペクトルが離散的な場合、ピッチにずれがある単音の場合でもスペクトルマッチングは非線形に行う必要がある。その理由は、図4のようにピッチのずれがある場合格子点上でマッチングを行う以上、最適な対応関係が階段状になるからである。特に低音に関しては、スペクトル成分の間隔は狭くなるにもかかわらずマッチング格子点間隔は一定なので、その影響は大きいはずである。

(iii) 計算量の問題

上述したように、線形伸縮を施してピッチのずれに適合したスペクトルテンプレートを用いた単純マッチングでは計算量が構成音数のべき乗に比例するが、DPマッチングによる計算量は、構成音数に依存しない。

以上から、ピッチにずれがある楽器の合奏音のスペクトルマッチングを行う場合は、スペクトルが周波数軸に非線形伸縮していると見なし、DPマッチングにより非線形な最適対応経路でマッチングをとる方が都合がよい。

ただし、DPマッチングを用いる場合、楽器音の調和性の拘束条件を考慮しないことになり、非線形の最適経路を探索することが悪影響を及ぼす可能性がある。従って、その影響の有無を確認するために、実験を通して検討する必要がある。

2.3 DPマッチングの定式化

DPマッチングを2つの周波数軸のスペクトル系列について適用する。隣り合う半音間のピッチの中間までの揺らぎを整合窓とすることで、ピッチの揺らぎに最大限に対応するマッチングとなる。入力音響信号、単・和音テンプレートそれぞれのス

ペクトル系列 A と B を、

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

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

とする。周波数軸変換関数を、

$$F = c(1), c(2), \dots, c(k), \dots, c(K) \quad (3)$$

とし、 $d(c(k))$ を $a_{i(k)}$ と $b_{j(k)}$ のベクトル間ユークリッド距離とすると、パターン A と B の累積距離 $D(A, B)$ を F を変化させたときの最小値として次式のように定義する。 $\omega(k)$ は非負の荷重係数である。

$$D(A, B) = \min_F \frac{\sum_{k=1}^K d(c(k)) \cdot \omega(k)}{\sum_{k=1}^K \omega(k)} \quad (4)$$

$c(k) = (i, j)$ と略記すると部分点列 $c(1), c(2), \dots, c(k)$ に対する部分積 $g(i, j)$ は、

$$g(i, j) = \min_{c(1), \dots, c(k)} \sum_{t=1}^k d(c(t)) \cdot \omega(t) \quad (5)$$

$$= \min_{c(k-1)} g(c(k-1)) + d(c(k)) \cdot \omega(k) \quad (6)$$

となる。対称形の $\omega(k)$

$$\omega(k) = (i(k) - i(k-1)) + (j(k) - j(k-1)) \quad (7)$$

を用いることにすると、図5より $c(k-1)$ は $(i-1, j)$ および $(i-1, j-1), (i, j-1)$ のいずれかである。よって、(6)式は次のように書ける。

$$g(i, j) = \min \left\{ \begin{array}{l} g(i-1, j) + d(i, j) \\ g(i-1, j-1) + 2d(i, j) \\ g(i, j-1) + d(i, j) \end{array} \right\} \quad (8)$$

これを、 $c(1) = (1, 1)$ から $c(K) = (I, J)$ まで順次求めていき、 A, B 両パターンの最小累積距離 $D(A, B)$ を得る。ここで、格子点 $(i(k), j(k))$ は整合窓条件は、

$$j(k) < 2^{\frac{1}{24}} \cdot i(k) \quad (9)$$

$$j(k) > 2^{-\frac{1}{24}} \cdot i(k) \quad (10)$$

を満たすものとする。半音上がるごとに基本周波数は $2^{\frac{1}{12}}$ 倍になるので(7), (8)式は、平均律で隣り合うピッチ間の中間までの範囲内を経路とすることができることを意味する。

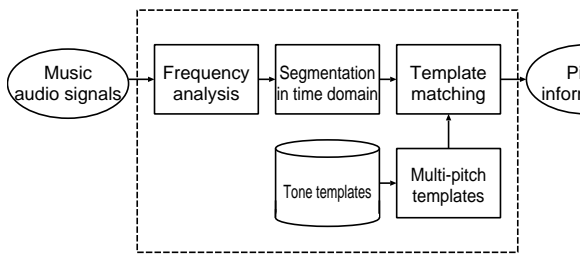


図 6 処理系の概要図

3. 楽器音の時間パターンの扱い

スペクトルマッチングを行う際、楽器音の音色の時間変動に対処する方法として以下のような処理系 (図 6) を構成した。その手順について以下に説明する。

3.1 セグメンテーション

セグメントは、音符長に相当し構成音のピッチが一定である区間として定義する。そこで、入力スペクトル系列から、音が立ち上がったと見なされる時点、または構成音が変化したと見なされる時点を実際の開始時点として扱う。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。 t をフレーム番号としてパワーの相対変化を表す尺度を $\alpha_1(t)$ 、周波数成分の変化を表す尺度を $\alpha_2(t)$ とし、それぞれを次の式として定義する。

$$\alpha_1(t) = \frac{P(t+1) - P(t)}{P(t)} \quad (11)$$

$$\alpha_2(t) = \frac{\sum_f (p(t+1, f) - p(t, f))^2}{P(t)} \quad (12)$$

ただし、 $P(t)$ はフレーム t における入力信号波形のパワー、 $p(t, f)$ はフレーム t 、周波数 f のスペクトル成分のパワー値を表す。そして、これらの尺度 $\alpha_1(t)$ 、 $\alpha_2(t)$ が以下の条件を満たすときの t を、セグメントの開始時点として扱う。ただし、 β_1 、 β_2 は実験で定められる定数である。

$$\alpha_1(t) > \beta_1 \quad \text{or} \quad \alpha_2(t) > \beta_2 \quad (13)$$

3.2 テンプレート

単音テンプレートは、曲ごとの最も低いピッチから最も高いピッチまでの半音間隔すべてのピッチのスペクトルである。なお、単音テンプレートは各楽器音のある 1 フレーム分のスペクトルとする。和音テンプレートは、最大 3 つまでの単音テンプレートを加算生成したものである。

3.3 マッチング方法

マッチングは、以下の手順で行う。

1. セグメント内すべてのフレームのスペクトルとテンプレートとの間で DP マッチングを行い、フレームごとに累積距離を求める。
2. セグメント内で累積距離をフレームごとに比較し、最小の累積距離 D_{min} を得る。
3. 1.2. の計算をすべてのテンプレートについて行う。
4. テンプレートごとに得られる D_{min} を比較し、 D_{min} が最小のテンプレートを認識結果とする。

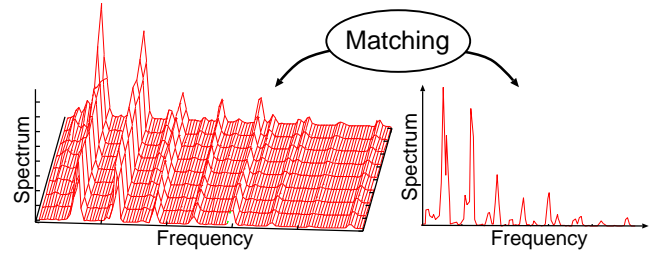


図 7 スペクトルテンプレートとセグメントとのマッチング

3.4 出力結果の補正

セグメンテーションにおいて、楽器音の立ち上がり時や消音時付近に余分なセグメントを設けてしまうことがある。そのようなセグメント内のスペクトル系列がどのテンプレートにも当てはまらない場合でも、その中で最適なものを推定結果として出力するため、出力推定結果にノイズが含まれることがある。そこでノイズの除去法として、以下のような HMM [16], [17] を用いた方法をとる。すべての単音または和音に対する隠れ状態を設定し、ノイズを含んだ推定ピッチ系列をフレームごとの状態遷移による出力シンボル系列と見なす。遷移確率はフレームごとに状態から状態に遷移する確率、状態出力確率は各状態において出力されるすべての単音または和音の尤度とし、このような HMM においてノイズを含んだ推定ピッチ系列が出力される尤度から状態遷移系列の推定を行う。その際、各状態に至るパスのうち最適なものだけを残し、残りのパスを枝刈りをするビタービ・アルゴリズム [16], [17] で効率良く最適パスを求めることができる。他のノイズ除去法としてメディアン・フィルタを用いる方法があるが [15]、HMM を用いるこの方法は、誤りの頻度 (オクターブ誤りが多い、など) を確率的なモデルに反映できるという利点をもつ。

4. 評価実験

4.1 実験系の概要

最大同時発音数が 3 の自然楽器 (ピアノまたはヴァイオリン) によって演奏された曲の音響信号を入力とし、それらのピッチを推定した。サンプリング周波数 44.1kHz、フレーム長 50ms (2048 点)、フレーム間隔 25ms (1024 点)、窓関数はハミング窓を用いて、入力音響信号の短時間スペクトル解析 (FFT) を行い、スペクトル系列を得た。ピアノによって演奏された曲に用いるテンプレートは C 2 から A 5 までの 46 個、ヴァイオリンによって演奏された曲に用いるテンプレートは G 2 から D 5 までの 32 個用意した。ただし各テンプレートは、ピアノ音の場合は音の立ち上がり時、ヴァイオリン音の場合は持続音中の適当な窓幅を切り出した際のスペクトルを正規化したものとした。また、(11) 式の β_1 と β_2 をそれぞれ、1.1, 0.007 とした。

4.2 認識率の計算

図 8 は、DP マッチングを用いたスペクトルマッチングによるピッチ推定結果である。図 9 は図 8 の出力結果をビタービ・アルゴリズムを用いた補正後の出力結果である。このように補

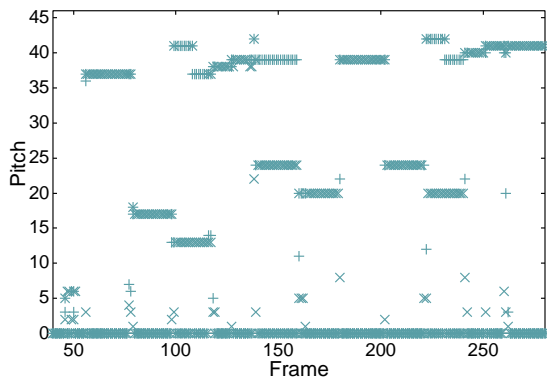


図 8 出力推定結果

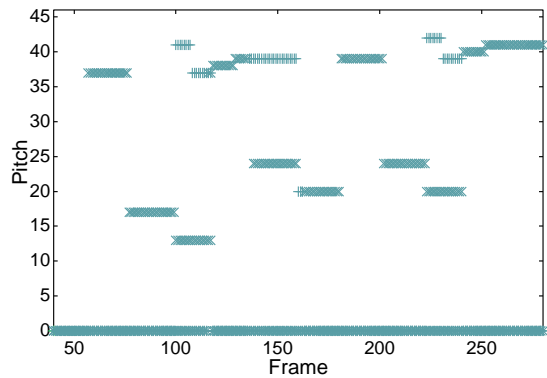


図 9 補正後の出力推定結果

正を行った出力推定結果から、認識率を算出した「音高の認識率」は曲の全音符数を分母とし、正しく出力された音符数の割合として定義する。また「和音の認識率」は和音の総数に対する和音正解（構成音の音高がすべての的中）の割合として定義する。ただし、以下の場合に音高の誤認識であると定義する。

1. 挿入 実際に演奏されたピッチはすべて出力されるが、余分にそれ以外のピッチも出力する場合
2. 脱落 実際に演奏されたピッチのうち一つでも出力されない場合
3. 置換 実際に演奏されたピッチが出力されず、代わりに異なるピッチを出力する場合

全音符数（和音の総数）を N とし、上記のいずれかの定義に当てはまる誤認識の総数（和音構成音の中で上記のいずれかの定義に当てはまるものがある和音の総数）を x とすると、音高および和音認識率 r を

$$r = \frac{N - x}{N} \times 100(\%) \quad (14)$$

とする。

4.3 実験結果

表 1 に DP マッチングと単純マッチングそれぞれのマッチング方法による各曲の和音の認識率を、表 2 に和音構成音の音高を推定した際の認識率を示す。結果から、特に弦楽器によって演奏されたものに関して、DP マッチングを用いる方がより高い認識率を得た。これは、弦楽器の演奏においてはピッチのずれが起りやすいからであると考えられる。また、表 3 に C4 以上のピッチを高音域、B3 以下のピッチを低音域とし、高音

域、低音域それぞれについての各手法の認識率を示す。同様に結果から、特に低音域に関して、DP マッチングを用いる方がより高い認識率を得た。すでに述べたように、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ため、非線形性の影響がより顕著になって表れたと考えることができる。また、すべての結果を通して DP マッチングを用いる方が高い性能が見られた。従って、楽器音の調和性という拘束条件を無視したことによる悪影響よりも非線形性を吸収したことの効果の方が大きかったと考えられる。

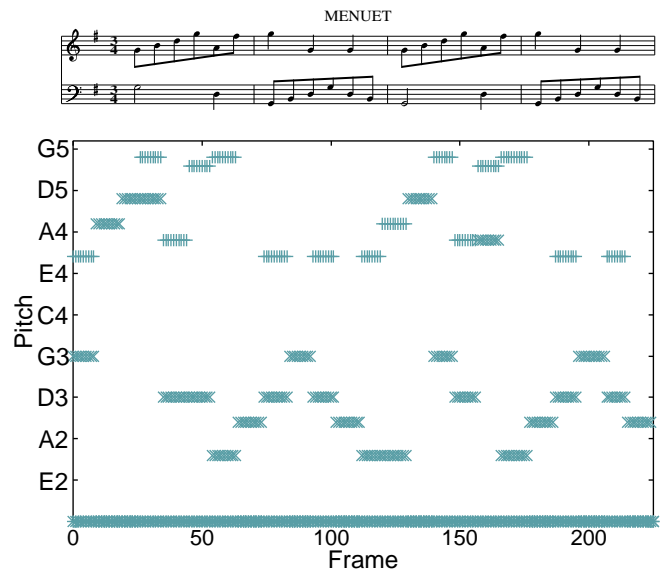


図 10 Bach: "Menuet" の一部の採譜結果 (ピアノロール表示)

表 1 和音の認識率

曲名	楽器	和音数	マッチング方法 (%)	
			DP	単純
Bach: "Menuet"	ピアノ	205	81.6	78.6
Mozart: "Menuet"	ピアノ	58	86.2	79.3
Pachelbel: "Kanon"	バイオリン	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

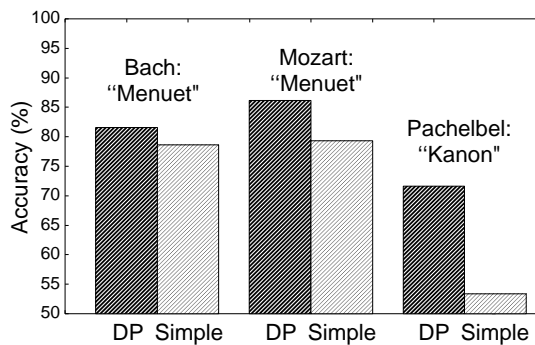


図 11 和音の認識率

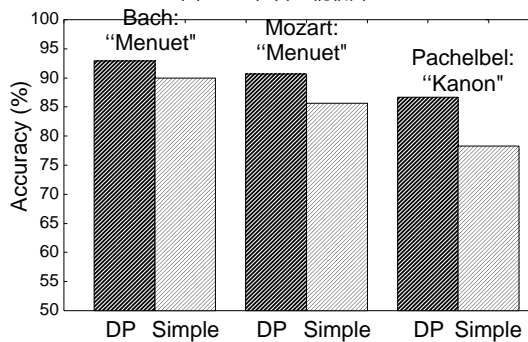


図 12 各音高の認識率

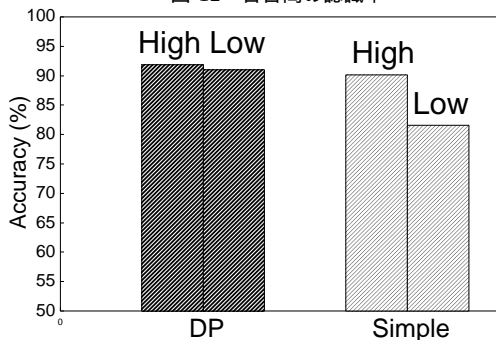


図 13 高・低音の認識率

5. む す び

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、周波数領域の 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。本研究の意義は、離散スペクトルを扱う上でこれまでに問題視されていなかったマッチングの非線形性の影響という仮説を考え、その解決策を見出そうとした点にある。

評価実験の結果、ピアノ 1 台による演奏曲での音高認識率は 92%、和音認識率は 82%であった。ヴァイオリンの 3 重奏での音高認識率は 87%、和音認識率は 71%であった。また、従来のマッチング方法と比較して、ピッチのずれの生じやすい弦楽器、および低音域において提案方法が有効に働いていることが分かった。

ピアノには非調和性という特有の性質があるが、その性質は楽器個体に依存する可能性があるため、本手法のように DP マッチングを用いる方が有利である。本実験では演奏曲とテンプレートに用いられた楽器は同一であったが、今後異なる個体間でその効果を検証したい。

DP マッチングを用いたスペクトルマッチングは、計算量が構成音数には依存しないが、二次元探索問題であるため計算量が大きく、計算効率をさらに向上させる必要がある。例えば、累積距離の計算をすべてのスペクトルテンプレートにわたって並行的に行い、計算過程で最小累積距離に比べてある程度大きい累積距離となったテンプレートの距離計算を中断していく方法 (ビームサーチ) の利用などが考えられる。

また、本処理系において、セグメント (入力スペクトル系列) とスペクトルテンプレートとの間でマッチングを行ったが、セグメンテーションの精度やスペクトルテンプレートの選び方として適当に楽器音波形から切り出したという点で、音色の時間変動に対する十分な対処ではなかった。そのため、今後音色の時間変動を吸収する方法について十分に検討していきたい。

6. 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。また、演奏データを提供された被験者の方々に感謝する。

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] 渡辺, 嶽間沢, 秋山: “器楽の自動採譜,” 電子情報通信学会情報・システムソサイエティ大会講演論文集, p. 325, 1998.
- [3] 渡辺, 嶽間沢, 秋山: “器楽の自動採譜 (その 2),” 電子情報通信学会総合大会, p.238, 1999.
- [4] 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.
- [5] 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.
- [6] 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誌, Vol. 33, No. 5, pp. 417-423, 1997.
- [7] 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.
- [8] 長束哲郎, 片寄晴弘, 井口征士: “異種楽器の音源分離システム,” 電子情報通信学会全国大会予稿集, D-5, 48, 1991.
- [9] 長束直樹, 才脇直樹, 井口征士: “異種楽器を対象とした採譜システム,” 電子情報通信学会'92 春大会予稿集, D-499, 1992.
- [10] 中村勲, 鈴木英男: “ピアノの発音機構と音色,” 日本音響学会誌, Vol. 49, No. 3, pp. 178-183, 1993.
- [11] 柏野, 中臺, 木下, 田中: “音楽情景分析の処理モデル OPTIMA における単音の認識,” 電子情報通信学会論文誌, Vol. J79-D-II, No. 11, pp. 1751-1761, 1996.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 新原, 今井, 井口: “歌唱の自動採譜,” 計測自動制御学会論文誌, Vol. 20, No. 10, 1984.
- [16] 古井貞照: “音声情報処理,” 森北出版, 1998.
- [17] 中川聖一: “確率モデルによる音声認識,” 電子情報通信学会, 1988.