

E-045

多重音の基本周波数評価尺度の最小値選択による実時間音高・音源数推定の試み

An experimental research of realtime multi-pitch estimation based on the minimum value selection of an evaluation function for multiple fundamental frequencies

宮坂広純† 松島俊明† 小田弘良†
Hirozumi Miyasaka Toshiaki Matsushima Hiroyoshi Oda

1. はじめに

筆者らは従来より尺八譜の情報処理システムの研究を行っており、機能の一部として自動譜めくりの実装を試みてきた[1]。自動譜めくり機能に要求される要素技術の1つに実時間での音高推定がある。尺八演奏だけに限れば単音を対象とすれば十分であるが、箏や三味線等などの合奏曲への応用を考慮すると、複数音への対応も必要となる。

近年、複数の楽音を含んだ音響信号から音高情報を抽出する研究が盛んに行われるようになり、例えば混合正規分布モデルのパラメータ推定[2]や混合正規分布の調波時間構造のクラスタリング[3]などの方法が提案されており、目覚ましい成果を挙げている。しかし、これらの方法では基本的にEMアルゴリズム等などによる繰り返し計算を用いているため、処理時間のコストが非常に大きく、自動譜めくりなど実時間処理が必要な用途に対してこれらの方法を直接適用することは困難である。本研究では、複素スペクトル内挿法[4]により求めた基本周波数候補と同時発音数の組み合わせを評価する新しい尺度を導入し、この評価尺度が最小となる推定音高の組み合わせを直接求めることで、音源数が未知な音響信号からの高速音高推定を可能としたので報告する。

2. 音高推定のための評価関数

多重音の音高推定を行うための問題の定式化方法はいくつかあるが、本研究では嵯峨山らにより考案されたハーモニック・クラスタリング[5,6]における定式化と同様の方法により定式化を行う。

音源に含まれる単音の基本周波数を Ω とすると、一般的に窓関数の影響等により観測されるスペクトルの形状は Ω を中心として左右に単調減少しかつ対象な形状となる。一般の楽音の場合は倍音構造を持つため、倍音の周波数が基本周波数 Ω の整数倍となることを仮定すると、 ω を周波数、 $f(\omega)$ を観測スペクトルとして、

$$D(\mu) = \sum_n \int_n \varphi(\omega, n\mu) \cdot f(\omega) d\omega \quad (1)$$

$$\text{ただし } T_n = \left\{ \omega \mid n = \arg \min_m |\omega - m\mu| \right\} \quad (2)$$

を最小とする μ として Ω は与えられる。ここで、 $\varphi(\omega, n\mu)$ は ω と μ の距離尺度を表す関数で、 $\omega = n\mu$ で極小値を取る左右対称な関数であれば何でも良いが、文献[5]と同様に

$$\varphi(\omega, n\mu) = (\omega - n\mu)^2 \quad (3)$$

に選ぶと、式(1)は

$$D(\mu) = \sum_n \int_n (\omega - n\mu)^2 \cdot f(\omega) d\omega \quad (4)$$

となる。また T_n は $(n-1)\mu$ と $n\mu$ の中間および $n\mu$ と $(n+1)\mu$ の中間を両端とする帯域を表している。

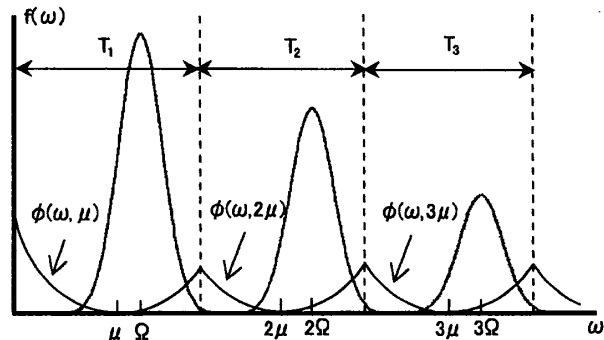


図1 単音の場合の基本周波数推定方法

音源数が複数の場合は、式(1)、式(2)を複数音に拡張した評価関数、即ち、音源数を K 、基本周波数の推定値を $\mu = \{\mu_1, \mu_2, \dots, \mu_K\}$ とすると、評価関数

$$D(\mu) = \sum_{k=1}^K \sum_n \int_n \varphi(\omega, n\mu_k) \cdot f(\omega) d\omega \quad (5)$$

$$\text{ただし } T_n^k = \left\{ \omega \mid (n, k) = \arg \min_{m, l} |\omega - m\mu_l| \right\} \quad (6)$$

を最小とする μ を求めれば良い。この最小値は、式(6)により与えられる周波数帯域 $T^k = T_1^k \cup T_2^k \cup \dots$ をクラスタと見なしてクラスタ重心である μ_k を

$$\mu_k = \frac{\sum_n \sum_{\omega_i \in T_n^k} \omega_i f(\omega_i)}{\sum_n \sum_{\omega_i \in T_n^k} f(\omega_i)} \quad (7)$$

により更新するという反復計算を行い、その収束値を求めることで得ることができるが、収束するまでの反復処理に時間がかかる、収束値が初期値に依存し、誤った値に収束することが多々あるため様々な初期値に対して試行する必要がある等の問題がある。そのため、クラスタリングによる解法を譜めくり等リアルタイム性が要求される用途に適用するのは現状では難しい。

† 東邦大学理学部情報科学科

そこで、本研究では、周波数ピークの検出精度に優れた複素スペクトル内挿法により検出されたピーク位置の中に正しい基本周波数が含まれている確率が非常に高い点に着目し、クラスタリングにより $D(\boldsymbol{\mu})$ の収束値を求めるのではなく、様々な組み合わせの $\boldsymbol{\mu}$ に対して直接 $D(\boldsymbol{\mu})$ の値を計算して $\boldsymbol{\mu}$ の推定を行うことにした。

3. 音高・音源数の同時推定アルゴリズム

まず、音源数(和音数) K が既知の場合の音高推定アルゴリズムを示す。短時間フーリエ変換(STFT)により得られたパワースペクトルのうち、一定以上のパワーを持つ上位 L 個(ただし $L > K$ とする)のピーク周波数 $\{\mu_1, \mu_2, \dots, \mu_L\}$ を複素スペクトル内挿法により抽出する。この L 個の基本周波数候補から選び出した K 個の基本周波数を小さい順に並べたベクトル

$$\boldsymbol{\mu} = (\mu_{j(1)}, \mu_{j(2)}, \dots, \mu_{j(K)}), \quad j(p) = 1, 2, \dots, L \quad (8)$$

に対して $D(\boldsymbol{\mu})$ を算出し、最小の $D(\boldsymbol{\mu})$ を与える $\boldsymbol{\mu}$ を基本周波数の推定値とする。従って音源数 K が既知の場合のアルゴリズムは以下ようになる。

アルゴリズム1: 音源数 K が既知の場合

- ① 複素スペクトル内挿法によりパワーのピーク位置を検出し、パワーの大きい順に L 個の基本周波数候補 $\{\mu_1, \mu_2, \dots, \mu_L\}$ を選出する。
- ② L 個の候補から K 個を選択して得られる ${}_L C_K$ 個の基本周波数の組み合わせの集合を $M = \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_{{}_L C_K}\}$ とする。
- ③ M の各要素について $D(\boldsymbol{\mu}_i)$ を算出し、最小値を取る $\boldsymbol{\mu}_i$ を基本周波数の推定値とする。即ち

$$\boldsymbol{\mu} = \arg \min_{\boldsymbol{\mu}_i} D(\boldsymbol{\mu}_i) \quad (\boldsymbol{\mu}_i \in M) \quad (9)$$

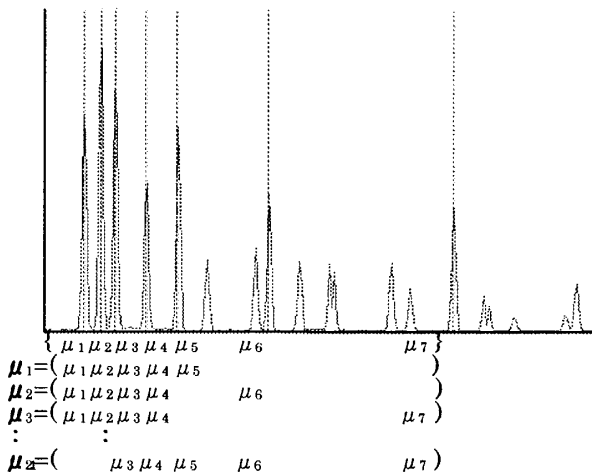


図2 複素スペクトル内挿法によるピーク検出と基本周波数候補の選択例($K=5, L=7$ の場合)

次に、音源数が未知の場合であるが、音源数が K の時の $\boldsymbol{\mu}$ を $\boldsymbol{\mu}^K$ と表すと、音源数が K の場合の式(5)の最小値 $\min D(\boldsymbol{\mu}^K)$ を $K=1$ から求めていき、最も小さい値を与える K および $\boldsymbol{\mu}$ に決定する方法が考えられる。しかし、評価関数の値 $D(\boldsymbol{\mu}^K)$ は、一般的に $\boldsymbol{\mu}$ の次元 K に対して単調に減少していくので、実際の音源数よりも大きい次元の $\boldsymbol{\mu}$ で最小となることが分かる。そこで、 $\boldsymbol{\mu}$ の次元数の増加に対するペナルティ係数 $p(K)$ を追加し、 $p(K)D(\boldsymbol{\mu}^K)$ を新たな評価関数とする。従って最大検出音源数(検出可能な最大の同時発音数)を N とすると、音源数が未知の場合の音高推定アルゴリズムは以下ようになる。

アルゴリズム2: 音源数 K が未知の場合

- ① 複素スペクトル内挿法によりパワーのピーク位置を検出する。
- ② $K=1$ とする。
- ③ パワーの大きい順に L 個 ($L > K$) の基本周波数候補 $\{\mu_1, \mu_2, \dots, \mu_L\}$ を選出する。
- ④ L 個の候補から K 個を選択して得られる ${}_L C_K$ 個の基本周波数の組み合わせの集合を $M^K = \{\boldsymbol{\mu}_i^K | i=1, 2, \dots, {}_L C_K\}$ とする。
- ⑤ M^K の各要素について $D(\boldsymbol{\mu}_i^K)$ を算出し、最小値を取る $\boldsymbol{\mu}_i^K$ を音源数が K と仮定した時の基本周波数の推定値とする。即ち

$$\boldsymbol{\mu}^K = \arg \min_{\boldsymbol{\mu}_i^K} D(\boldsymbol{\mu}_i^K) \quad (\boldsymbol{\mu}_i^K \in M^K) \quad (10)$$

- ⑥ $K < N$ ならば $K=K+1$ として③へ戻る。
- ⑦ $p(K)D(\boldsymbol{\mu}^K)$ を最小とする $\boldsymbol{\mu}^K$ を音源数 K と基本周波数 $\boldsymbol{\mu}$ の推定値とする。即ち

$$(\boldsymbol{\mu}, K) = \arg \min_{\boldsymbol{\mu}^K} p(K)D(\boldsymbol{\mu}^K) \quad (K=1, 2, \dots, N) \quad (11)$$

4. 実験結果

本方式による実験結果を以下に示す。実験には Pentium4 3.2GHz CPU, 1.00GByte RAM を搭載した Windows XP マシンを、STFT の窓長は 4096, 窓関数はハミング窓を用いた。

まず、本方式の基本的な性能を評価するために、MIDI 音源により作成した 3 和音(CEG), (DFA), (EGB), (FAC), (GBD), (CDE), 5 和音(CEGBD), (DFACE), (EGBDF), (FACEG), (GBDEA), (ACEGB), (BDFAC) の音響データに対して、アルゴリズム1により音高推定を行った結果を図3に示す(実験1)。ただし基本周波数候補数 $L = K+2$ とした。この結果より、音源数が既知の場合は、アルゴリズム1によりほぼ正確に複数音の音高推定が可能であることが分かる。

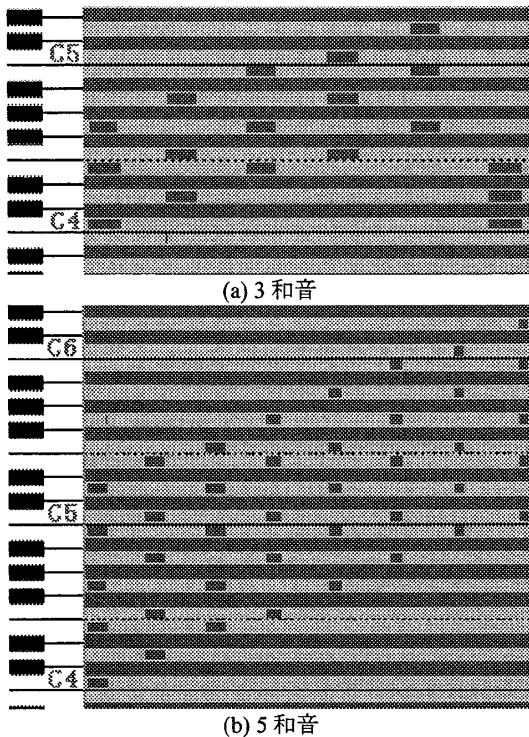


図3 実験1：音源数が既知の場合の音高推定結果

次に、音源数が変化する場合について、アルゴリズム2により音高推定を行った結果を図4に示す(実験2)。ただし、補助実験の結果から、ペナルティ係数 $p(K) = K$ 、基本周波数候補数 $L = K + 2$ に設定した。また最大検出音源数 $N = 6$ として実験を行った。音源数が4を超える部分では処理落ちが生じることがあったが、図4からも分かるように、アルゴリズム2による音高推定方法は音源数が未知の場合に有効であることが分かる。

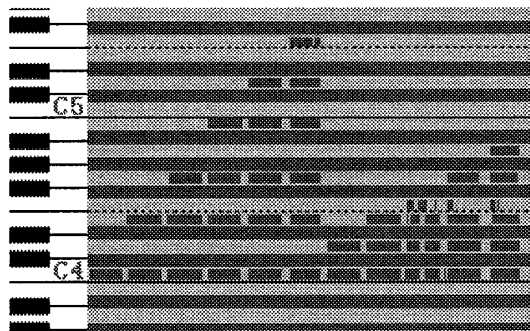


図4 実験2：音源数が未知の場合の音高推定結果

最後に、RWC 研究用音楽データベース[7]に含まれている楽曲サンプル(MDB-J-2001 No.9)に対してアルゴリズム2を適用した結果を図5に示す(実験3)。 $N = 4$, $L = K + 2$ とし、 $p(K)$, $\phi(\omega, n\mu)$ は実験2と同一である。最低音のD3の検出ミスや4和音部分での検出漏れが散見されるが、リアルタイムの処理にもかかわらず音源数および音高の変化に比較的良く対応できており、本方式が実際の楽曲に対しても十分適用可能であることが分かった。

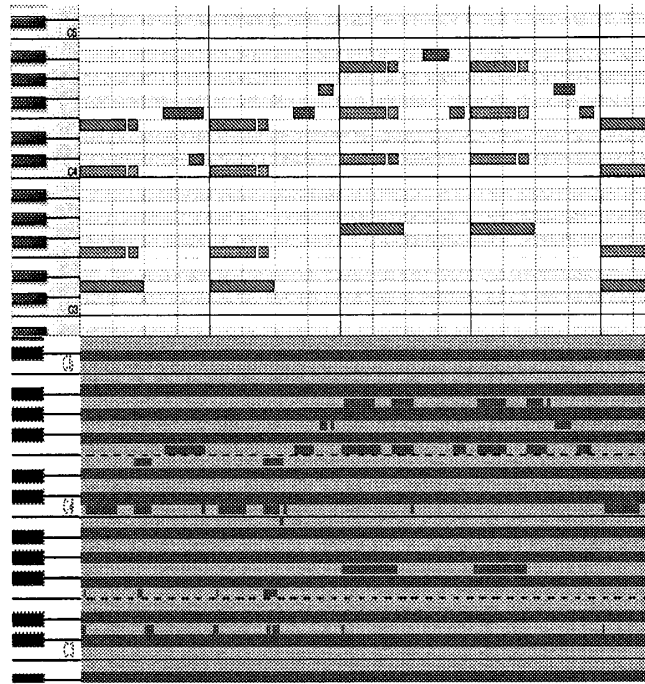


図5 実験3：楽曲サンプルデータに対する音高推定結果 (上段：正解MIDIデータ，下段：音高推定結果)

5. まとめ

以上のように、本方式では、比較的簡単な処理により高速かつ高精度に音高を推定することが可能となった。

一般に音源数の推定は難しく、今回は経験的にペナルティ係数を $p(K) = K$ としたが、この係数の設定には今後十分な検討が必要である。

また、低音部やオクターブ重なりなどの和音の検出精度が良くない点、最大音源数が増える(概ね5音以上)と処理量が増大し、実験機では処理落ちが生じることがある等の問題点が残されているが、4音程度までの和音をほぼ正確に検出することが可能であれば、既知の楽譜中の演奏箇所の推定は可能と思われるので、譜めくりシステムでの利用に十分適用可能な方法と考えている。

参考文献

- [1]松島俊明, "尺八譜のマルチメディア情報処理", 画像電子学会第34回年次大会予稿集, pp.23-30 (2006)
- [2]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", 電子情報通信学会論文誌, Vol.J84-D-II, No.1, pp.12-22 (2001)
- [3]亀岡弘和, 西本卓也, 嵯峨山茂樹, "調波時間構造クラスタリング(HTC)による音楽音響特徴量の同時推定", 情報処理学会研究報告, 2005-MUS-61-12, pp.71-78 (2005)
- [4]原裕一郎, 井口征士, "複素スペクトルを用いた周波数同定", 計測自動制御学会論文集, Vol.19, No.9, pp.718-723 (1983)
- [5]亀岡弘和, 西本卓也, 篠田浩一, 嵯峨山茂樹, "ハーモニッククラスタリングによる多重音の基本周波数推定アルゴリズム", 情報処理学会研究報告, 2003-MUS-50-5, pp.27-32 (2003)
- [6]亀岡弘和, 西本卓也, 嵯峨山茂樹, "ハーモニッククラスタリングと情報量基準による音楽の音高/音源数の推定", 情報処理学会研究報告, 2005-MUS-62-5, pp.27-32 (2005)
- [7]後藤真孝, 橋口博樹, 西村拓一, 岡隆一, "RWC 研究者用音楽データベース: クラシック音楽データベースとジャズ音楽データベース", 情報処理学会研究報告, 2002-MUS-44-5, pp.25-32 (2002)